

CONTROL SYSTEM THEORY

• Feedback Engineering

GLADWYN LAGO

LLOYD M. BENNINGFIELD

PETER L. BALISE

MECH. ENG. DEPT.

UNIV. OF WASH.

25.12

BY G. V. LAGO AND D. L. WAIDELICH

Transients in Electrical Circuits

CONTROL SYSTEM THEORY

Feedback Engineering

GLADWYN LAGO

UNIVERSITY OF MISSOURI

and

LLOYD M. BENNINGFIELD

UNIVERSITY OF MISSOURI

THE RONALD PRESS COMPANY • NEW YORK

Copyright © 1962 by
THE RONALD PRESS COMPANY

All Rights Reserved

No part of this book may be reproduced
in any form without permission in writing from
the publisher

Library of Congress Catalog Card Number: 62-9759

PRINTED IN THE UNITED STATES OF AMERICA

To Our Seniors

PREFACE

Automatic control system theory cuts across all fields of engineering and in so doing creates a new field of its own. Because of the broad nature of the control problem and the diversified character of what is to be controlled, the theory is an assemblage of many branches of science and technology, with such diverse applications as positioning of mechanical objects, controlling fluid flow, viscosity, temperature, humidity, pressure, and chemical reactions, handling of parts, assembly of parts, and data processing.

The curricula of almost all engineering colleges include courses in control system theory taught in the Departments of Chemical Engineering, Electrical Engineering, Industrial Engineering, and Mechanical Engineering, among others. Therefore, this book is written from as general a point of view as possible. Although it is directed specifically to the needs of electrical engineering students, it also features chapters on mechanical and electromechanical systems, hydraulic systems, and pneumatic systems. It is assumed that the student has been introduced to transient analysis at a level comparable to that represented by Lago and Waidelich's *Transients in Electrical Circuits*.

This book is the outgrowth of two upper-level courses taught at the University of Missouri. The first is a required course taught in the first semester of the senior year; the second is an elective for seniors and graduate students. These courses give the student a broad perspective of the analysis and design of control systems, from which he may, if he chooses, proceed to more advanced study. Experience at the University of Missouri has clearly demonstrated the value of using the pole-zero approach consistently, of stressing the root-locus and the inverse root-locus methods as equally important counterparts, and of including some aspects of network synthesis to introduce the direct synthesis of control systems.

The first two chapters introduce the basic concepts and nomenclature of closed-loop systems, and the Laplace transformation and several special techniques for solving integrodifferential equations. There follow three parallel discussions of systems: electric circuits (including the concepts of transfer functions and block diagrams), mechanical and electromechanical systems, and hydraulic and pneumatic systems.

Transient response and steady-state sinusoidal response are treated with equal importance in control system analysis, and their relationship is studied in detail. Block-diagram transformations are developed as the basis for the analysis, and the pole-zero approach unifies it. Three chapters are devoted to the use of Nyquist diagrams and Bode plots in studying analysis and design by the steady-state sinusoidal response. The root-locus method is developed as a means of studying the movement of the poles of the closed-loop system as the gain of the open-loop system is varied. The inverse root-locus method also is developed.

The final portion of the book treats network synthesis and the direct method of synthesis of control systems, and analog computer theory with emphasis on the application to control.

The authors acknowledge with sincere thanks the helpful suggestions and encouragement received from many students and colleagues at the University of Missouri and elsewhere. The authors are indebted to Mr. Bert Gastineau and Dr. John N. Warfield, who prepared the table of Laplace transforms. Gratitude is due also to Messrs. O. Bodenstein, Luc Baerman, R. M. Scott, John Scobee, Terry Stair, and Charles Couranz for their help in preparing classroom notes and original drawings for the illustrations, and to Mrs. Sheila Lipsey, Miss Linda Hayward, and Mrs. Laverne Wiggs, who so painstakingly did the typing.

GLADWYN LAGO
LLOYD M. BENNINGFIELD

Columbia, Missouri
March, 1962

CONTENTS

CHAPTER	PAGE
List of Symbols	ix
1 Control System Concepts	3
2 The Laplace Transformation	16
3 Electric Circuits	52
4 Mechanical and Electromechanical Systems	88
5 Hydraulic and Pneumatic Systems	128
6 Analysis of Simple Control Systems	152
7 Stability	204
8 Design Using Nyquist Diagrams	252
9 Frequency Response Characteristics	282
10 Transient and Steady-State Relationships	315
11 Root-Locus Methods	352
12 Network Synthesis	397
13 Direct Synthesis Methods	436
14 Analogs and Analog Computers	457
15 Summary	491
Appendix: Laplace Transforms	496
Index	501

LIST OF SYMBOLS

- $A_1, \dots, A_i$ Piston area; points in $1 + GH$ plane; general angles.
 $A(j\omega)$ Steady-state amplification ratio of an amplifier.
 B Magnetic flux density; hydraulic fluid bulk modulus.
 $B(s), B(j\omega)$ Transforms of feedback signals.
 $C_0, C_1, \dots, C_i$ Constants; electrical capacitance; polynomial coefficients.
 C_h, C_p Hydraulic and pneumatic capacitance.
 $C(s), C(j\omega)$ Transforms of controlled output signals.
 $D_1, \dots, D_i, D_{\text{eff}}$ Viscous friction coefficients; diameter.
 D_G, D_H Denominator polynomials from $G(s)$ and $H(s)$.
 $D(s)$ General denominator polynomial in s .
 $E_1, \dots, E_i$ Voltage; complementary solutions for differential equation.
 E_p Particular solution for differential equation.
 $E(s), E(j\omega)$ Transforms of error or actuating signals; voltage transforms.
 $F_1, \dots, F_i$ Force.
 $F_i(j\omega)$ Phasor or steady-state transform of $f_i(t)$.
 $F_i(s)$ Laplace transform of general $f_i(t)$; general function of a complex variable.
 $G_1(s), \dots, G_i(s)$ General transfer functions (open-loop forward transfer functions in particular).
 $G_i(j\omega)$ Steady-state transfer functions.
 H Magnetic field intensity; constant.
 $H_1(s), \dots, H_i(s)$ Transform of fluid head.
 $H(s), H(j\omega)$ Feedback transfer functions.
 $I_A, I_1, \dots, I_i$ Current.
 I_m Maximum value of phasor $I(j\omega)$.
 $I(s), I(j\omega)$ Transforms of general input signals; transforms of current.
 $J_1, \dots, J_i, J_{\text{eff}}$ Mass moment of inertia.
 $K, K_1, \dots, K_a, \dots, K_i$ General constants; static error coefficients.
 $K_{s1}, \dots, K_{si}, K_{s \text{ eff}}$ Spring constants.
 L_i Inductance; contours in the complex $GH(s)$ plane.
 $M_1, \dots, M_i$ Mass; magnitude of closed-loop frequency response; manipulated variable transforms.
 M_m Maximum magnitude of closed-loop frequency response.
 N Designation of the constant angle $C(j\omega)/R(j\omega)$ contours; number of revolutions of Nyquist plot about -1 point in $G(s)H(s)$ -plane.
 N_G, N_H Numerator polynomials of $G(s)$ and $H(s)$.
 $N(s)$ General numerator polynomial in s .
 $O(s)$ Transform of general output signals.
 P Number of open-loop poles in right-hand half s -plane in applying Nyquist criterion.

- $P_1(s), \dots, P_i(s)$ Transform of pressure.
 Q Total volume of flow.
 $Q_1(s), \dots, Q_i(s)$ Transform of volume flow rates.
 $R, R(s), R(j\omega)$ Transforms of reference input signals.
 $R_1, \dots, R_i, R_h, R_p$ Electrical, hydraulic, and pneumatic resistance.
 R_{gas} Universal gas constant.
 T Time delay; period of periodic functions.
 $T, T_1, \dots, T_i$ Torque; temperature.
 T_d Delay time of step response.
 T_p Time to peak of step response.
 T_r Rise time of step response.
 T_s Settling time of step response.
 $T(j\omega), T(s), T'(s)$ Closed-loop transforms, C/R .
 $U, U(\sigma, \omega)$ Real part of a function of a complex variable $s = \sigma + j\omega$.
 $V, V(\sigma, \omega)$ Imaginary part of a function of a complex variable $s = \sigma + j\omega$.
 $V_1, \dots, V_A, \dots, V_i$ Voltage.
 $V(s)$ Transform of velocity.
 $W, W_1, \dots, W_i$ Weight; work.
 W_{total} Total weight of flow of air.
 $X(j\omega), X(s), X_1(s), \dots, X_i(s)$ Transform of linear displacement.
 $Y(j\omega), Y(s), Y_1(s), \dots, Y_i(s)$ Electrical admittance.
 Z Number of zeros of system characteristic polynomial in the right-hand half s -plane as determined from the Nyquist criterion.
 $Z(j\omega), Z(s), Z_1(s), \dots, Z_i(s)$ Electrical impedance.
 $\mathcal{S}[\]$ Direct Laplace transform.
 $\mathcal{S}^{-1}[\]$ Inverse Laplace transform.
 $\text{Re}[\]$ Real part of $[\]$.
 $\text{Im}[\]$ Imaginary part of $[\]$.
 $a, a_0, a_1, \dots, a_i$ General coefficients; constants; points in the s -plane; transformer turns ratio.
 $b, b_0, b_1, \dots, b_i$ General coefficients; constants; time-varying feedback signal; transformer turns ratio.
 $c, c(t)$ Time-varying controlled output signal.
 $c_1, \dots, c(t)$ General coefficients; constants.
 c_m Motor shaft position.
 $c(s)$ General polynomial in s .
 $d_1, \dots, d_i$ Distance in s -plane.
 $e, e(t)$ Time-varying error or actuating signal.
 $e_0, e_1, \dots, e_i$ Voltage.
 e_{asympt} Value of $e(t)$ approached asymptotically with time.
 $f_i(t)$ General function of time.
 $f_i^{(n)}(t)$ n th derivative of $f_i(t)$ with respect to time.
 $f_v, f_{v1}, \dots, f_{v1}$ Coefficient of translational viscous friction.
 g Acceleration of gravity.
 g_m Vacuum tube transconductance.
 $g(t)$ General function of time.
 $h, h_1, \dots, h_i$ Fluid head.
 $i, i_1, \dots, i_i$ Current.
 j $\sqrt{-1}$

- k Ratio of lead and lag compensation pole-zero pair; negative of normalized closed-loop zeros.
 k_{-i} Residues at the poles of a function of s .
 $l, l_1, \dots, l_i$ Contours in the s -plane; physical l lengths.
 m Negative of normalized closed-loop poles.
 $m(s)$ Even part of a polynomial $p(s)$.
 $m(t)$ Time-varying manipulated signal.
 $n, n_1, \dots, n_i$ Gear ratios; polytropic process exponent.
 $n(s)$ Odd part of a polynomial $p(s)$.
 p Normalized variable s .
 $p, p_1, \dots, p_i$ Negative of poles in the s -plane; pressure.
 $p(s)$ General polynomial in s .
 $q, q_1, \dots, q_i$ Volume flow rates.
 $q(s)$ General polynomial in s .
 $r, r(t)$ Time-varying reference input signals.
 $r_1, \dots, r_i$ Gear radii.
 r_p Incremental vacuum tube plate resistance.
 s Complex frequency variable of the Laplace transform.
 $s_1, \dots, s_i$ Specific values of s .
 t, t' Time variables.
 $t(t)$ Time-varying torque, time-varying closed-loop impulse response.
 u General variable.
 $u_0(t)$ Unit-impulse function.
 $u_{-1}(t)$ Unit-step function.
 $u_{+1}(t)$ Unit-doublet function.
 $u_{-2}(t)$ Unit-ramp function.
 v General variable.
 $v_1, \dots, v_i, v(t)$ Time-varying velocity; specific volume; voltage.
 $v(s)$ General function of s .
 $w_1, \dots, w_i$ Weight flow rates.
 $x, x_1, \dots, x_i$ Linear displacement variable; general variable; real part of normalized variable p .
 x_{tran} Transient part of displacement.
 x_{ss} Steady-state part of displacement.
 y General variable; imaginary part of normalized variable p .
 $y_A, \dots, y_{11}, \dots, y_{ij}$ Network short-circuit admittances.
 $z, z_1, \dots, z_i$ General complex variable; negative of s -plane zeros; reciprocal of s .
 $z_A, \dots, z_{11}, \dots, z_{ij}$ Network open-circuit impedances.
 $\Delta(), ()_{\Delta}$ Increments in $()$.
 $\Theta(s)$ Transform of angular position.
 $\Omega(s)$ Transform of angular velocity.
 $\alpha, \alpha_1, \dots, \alpha_i$ Angular position; general angles; real part of complex s -plane pole; ratio of poles and zeros in compensation functions; negative of real s -plane poles; network division factor in parallel ladder synthesis.
 β Imaginary part of complex s -plane poles; negative of real s -plane poles; network division factor in parallel ladder synthesis.
 γ_i Negative of closed-loop poles.
 δ Pulse width; general angles.
 ϵ Small quantity.

- ζ Damping ratio of complex s -plane pole pairs.
 η Percent of maximum hydraulic pump tilt-plate displacement.
 $\theta, \theta_1, \dots, \theta_i$ Angular displacement.
 μ Vacuum tube amplification factor; absolute fluid viscosity.
 ϱ Specific weight; radius to a point in s -plane.
 σ Real part of the Laplace variable s .
 σ_0 Abscissa of absolute convergence.
 $\sigma_1, \dots, \sigma_i$ Negative of close-loop zeros.
 τ Variable of integration in convolution integral.
 $\varphi_1, \dots, \varphi_i$ General angles; magnetic flux.
 ψ General angles, magnetic flux linkage.
 $\psi(s)$ Reactance function.
 ω Imaginary part of the Laplace variable s ; real frequency.
 ω_d Damped angular velocity.
 ω_m Frequency of maximum closed-loop gain.
 ω_n Undamped natural angular velocity.

CONTROL SYSTEM THEORY

CONTROL SYSTEM CONCEPTS

1-1. Introduction. Since the beginning of recorded time, man has sought to make his life easier by producing devices that would relieve him of his daily drudgery. The industrial revolution in the nineteenth century relieved man's muscles of the burdens of many tasks by using energy obtained from coal, oil, and the flow of water.

The development of automatic control systems during the past few years has been referred to by some as the second industrial revolution. The first industrial revolution relieved man's muscles whereas the second industrial revolution relieves man's brain as well as his muscles.

Automatic control cuts across all fields and in the process creates a new field of its own. Automatic control encompasses such things as positioning of mechanical objects; controlling fluid flow, viscosity, temperature, humidity, pressure, and chemical reactions; handling parts; assembly of parts; data processing; and inventory control.

Because of the broad nature of the control problem and the diversified character of the things that are to be controlled, the theory of control systems is a bringing together of many branches of science and technology.

1-2. Open-Loop and Closed-Loop Control Systems. Control systems are generally divided into two categories; open-loop (open-cycle) and closed-loop (closed-cycle). The distinction between the two terms can best be made by the use of several examples.

Traffic lights that are controlled by a timing device are an example of an open-loop system. The light sequence of green-amber-red continues on a timing basis without any regard to the nature of the thing that is being controlled, namely, the flow of traffic. Everyone can recall instances in which a traffic light turned red and halted a flow of traffic when there were no cars on the side street. In contrast to this, a policeman in control of traffic is a part of a closed-loop system. That is, the policeman adjusts his timing of traffic flow as dictated by the nature of the traffic flow itself. For example, the policeman will not stop the flow of traffic on one street unless there are cars waiting on the side street.

Another example of an open-loop system is that of the automatic washing machine. The machine goes through its sequence of operations of soak, wash, rinse, and dry based on its timing device. The machine will wash each load in exactly the same way regardless of how dirty any particular load may be. As a matter of fact, the machine, once started,

goes through the same set of operations even if no clothes are in the machine. The desired result of the washing process is a certain degree of cleanliness of the clothes. The automatic washing machine is an open-loop system because the cleanliness of the clothes being washed in no way changes the washing cycle of the machine. In contrast to this, a laundress will scrub more vigorously on a dirty garment than she will on one that is almost clean. Therefore, she is a part of a closed-loop system.

With these two examples to clarify the concepts, certain of the meanings of the terms open-loop and closed-loop are evident. In an open-loop system the input is applied directly to the system and the output behaves in accordance with the characteristics of the system. In a closed-loop system the output is compared with the input and the difference between the two quantities is used to control the system.

A third example is that of a house heated by a gas furnace. This can be either an open-loop or a closed-loop system, depending upon how the furnace is controlled. In an open-loop system, the input valve controlling the flow of gas to the furnace is given a certain setting. The output of the system is the temperature of the house and this output reaches a certain value that is determined by the outside temperature, the outside wind velocity, the prevalence of sunshine, the pressure in the gas mains, and thermal characteristics of the house along with many other factors.

In a closed-loop system, the temperature of the house itself determines the amount of gas supplied to the furnace by means of some type of thermostat with a desired setting. If the temperature of the house falls below this setting, the furnace operates until the temperature reaches the desired value. The loop is closed in the sense that the temperature of the house controls the thermostat, which in turn operates the furnace, thus controlling the temperature.

Some of the advantages of a closed-loop system as compared with an open-loop system can be deduced from this example. A person controlling the temperature of a house by an open-loop system can consult the weather forecast each morning and set the valve controlling the flow of gas to the furnace at some appropriate value. If the weather deviates suddenly from the prediction by becoming much colder, the temperature of the house will fall far below the desired value. In contrast to this, the thermostat in a closed-loop system is set at a certain value. If the weather is mild, the furnace runs seldom. If the outside temperature falls, the thermostat simply turns the furnace on for longer periods of time, thus maintaining the desired house temperature.

1-3. Block Diagram. The basic operation of an open-loop and a closed-loop system can be shown by the respective block diagrams of Fig. 1-3-1. In Fig. 1-3-1a, the input of the open-loop system feeds into a power-amplifier stage, which in turn drives the controlled quantity. A disturbance is an undesired signal which can also affect the controlled quantity. An example of such a disturbance was given in the above open-

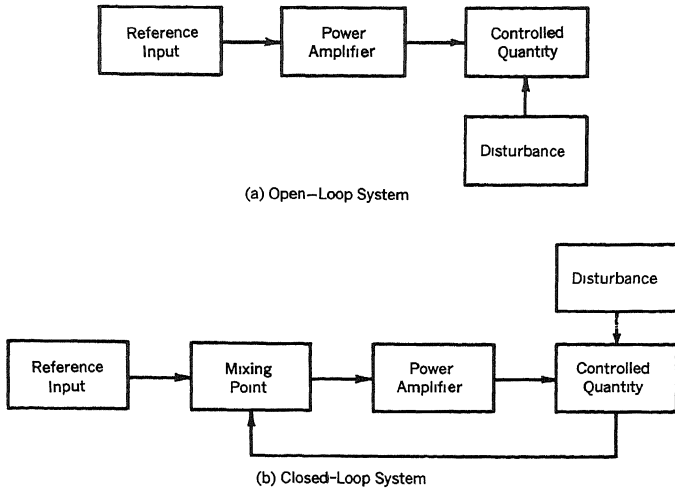


Fig. 1-3-1.

loop example, in which the outside temperature fell by an unexpected amount, causing the house temperature to fall below the desired value.

The input of the closed-loop system feeds into a mixing point along with a signal coming back from the output, which is the controlled quantity. The signal out of this mixing point is the difference between the input and output signals. It is called the actuating signal because it is this signal that actuates the power-amplifier stage. The power amplifier in turn drives the controlled quantity. A disturbance also can affect the controlled quantity. However, because of the nature of the closed loop, where the controlled quantity is affected by the disturbance, the output signal changes the actuating signal, which makes a correction (or at least a partial correction) in the controlled quantity.

In the simple system of Fig. 1-3-1b, the actuating signal can also be called the error signal. This term implies that ideally it is desired to have the controlled quantity exactly follow the commands of the reference input. If the output does not take on this desired value, it is in error by an amount equal to the difference between its desired value and its actual value. In this system the error is equal to the actuating signal. In a more general system there may be elements in the feedback path, that is the path from the controlled quantity back to the mixing point. If there are elements in the feedback path, the error and the actuating signal are not identical.

At first glance it may seem that the designer of a control system is in an impossible situation: he wishes the error to be zero, but at the same time the closed-loop system does not function unless there is an error. The answer to this enigma is that nothing in the world is exact and things can be done only within certain tolerances. If the error of operation of a closed-loop system can be kept within specified limits, the operation of the system is satisfactory.

The concept of comparing the output of a system with the input and letting the difference be the quantity that actuates the system may at first thought seem novel. However, a little reflection reveals that much of nature and human behavior follows such a pattern. For example, a person goes to the faucet to fill a glass with water. He does not make mental calculations as to water pressure and the volume of water flow for a given position of the handle, then turn it for a precalculated amount of time. Instead, he watches the level of water in the glass, comparing it with the desired level and controlling the rate of water flow to bring this error to zero.

As a second example, the housewife threading a needle does not look at the thread and the eye of the needle, then gaze out of the window while her hands follow a predetermined path. Instead, she compares the position of the thread and the position of the needle's eye, and moves her hand in such a way as to reduce the distance between these positions to zero.

Another favorite example is the driving of a car. The driver does not look far ahead and precalculate his every turn. Instead he actually compares the location of the car on the road with the desired location and turns the steering wheel to reduce the error to zero.

As a last example, a situation in human relations is given. A man tells his wife he will be home for dinner at 6:00. If there is an error in his timing and he arrives home much later than he promised, the wife applies a corrective force to the husband's actions. The later the hour of his arrival, the greater the corrective force.

It can be seen that the concepts coming from control system theory apply to human relations as well as to things. The term "cybernetics" has been used by Wiener¹ to convey the general meaning, as in the title of his book, *Cybernetics—Control and Communication in the Animal and the Machine*. Although the theories discussed here are applicable to the animal, this book is mostly concerned with the machine.

1-4. Illustrations of Control Systems. Examples of several types of control systems follow.

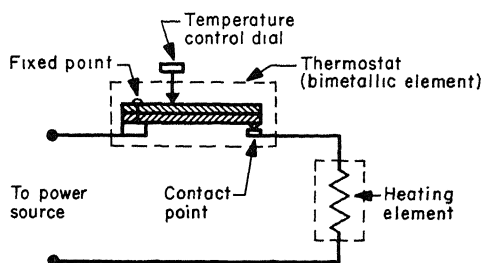


Fig. 1-4-1.

TEMPERATURE CONTROL. Fig. 1-4-1 is a schematic diagram of an electric iron. The dial on the iron is set for the proper temperature for ironing cotton, silk, rayon, etc., and the iron is then plugged into the power source. The thermostat is closed and the heating element is connected to the power source. The temperature of the sole plate increases until it reaches the temperature demanded by the dial setting. After this, the thermostat opens and the heating element is disconnected from the power source.

When this system is compared with the block diagram of Fig. 1-3-1b, the following functional relationship can be deduced. The reference input is the desired temperature of the sole plate as represented by the dial setting. The mixing point (error detector) is the thermostat which compares the temperature of the sole plate with that demanded by the dial setting. The input to the power amplifier might be thought of as the power required to operate the thermostat, and the output of the power amplifier as the power delivered to the heating element. The controlled quantity is the temperature of the sole plate. The disturbance is any influence that comes from outside the system and changes the temperature of the sole plate, such as ironing a damp garment.

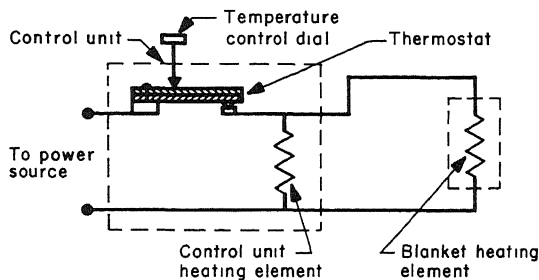


Fig. 1-4-2.

Fig. 1-4-2 is a schematic diagram of one type of circuit that can be used for an electric blanket. The portion of the circuit shown inside the dotted box is the control unit. The method of operation of the control unit alone is essentially that of the electric iron just discussed. This portion of the system is closed-loop in that the heating element keeps the temperature inside the control unit at a desired temperature as set on the dial. This closed-loop system in turn operates the blanket heating element as an open-loop device. If the blanket and control unit are in the same room (as they should be) and the temperature falls, the heating element inside the control unit operates for longer periods of time to maintain the temperature inside the control unit. The blanket heating element in turn operates for longer periods of time.

The problem in this type of system is to design the open-loop system so that the blanket is heated properly. A more desirable (and also more

costly) system would be to close the loop so that the temperature of the blanket itself is controlled instead of that of the control unit. This can be done by incorporating a control element in the blanket along with the heating element. The resistance of the control element varies with temperature. If the resistance is below a desired value, the heating element operates to increase the resistance. In this way the temperature of the blanket is maintained.

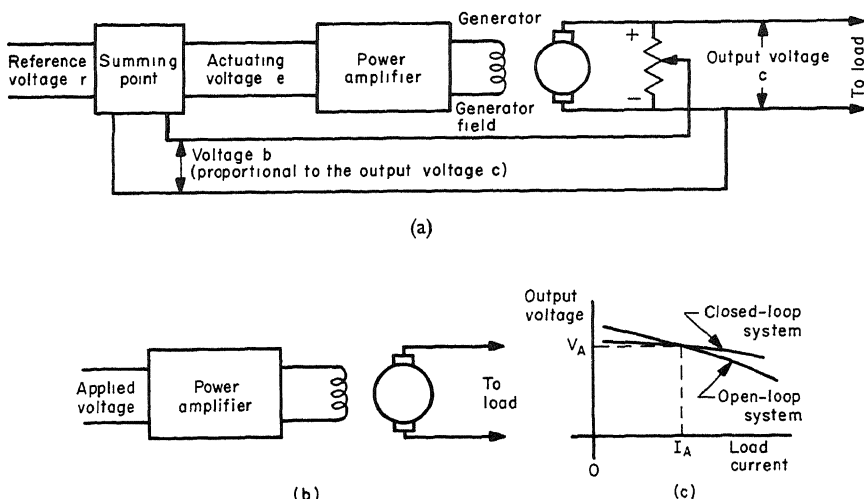


Fig. 1-4-3.

VOLTAGE REGULATION. Fig. 1-4-3a is a schematic diagram of a voltage regulator. It is of interest first to investigate the system when it is operated as an open-loop device. The open-loop system is shown schematically in Fig. 1-4-3b.

The generator is designed to deliver an output voltage V_A at a load current I_A . In the open-loop system this means that the proper voltage must be applied to the input of the power amplifier so that these conditions are met. With this voltage applied to the input, the load current is varied and, in accordance with the regulation of the system, a curve such as that shown in Fig. 1-4-3c is obtained.

In contrast to this, the action of the closed-loop system is examined. In this case, the actuating voltage e is the difference between the reference voltage r and voltage b , as shown by the equation

$$e = r - b \quad (1-4-1)$$

The system is adjusted to deliver V_A volts at I_A load current to coincide with the open-loop system. However, when the load current increases in

the closed-loop system, the output voltage becomes smaller, voltages c and b become smaller, and e becomes larger as can be seen from Eq. 1-4-1. The larger the actuating voltage e the larger the output of the generator, to the end that much of the loss of the output voltage is recovered. Therefore, the closed-loop system has a much smaller voltage regulation than the open-loop system, as is shown in Fig. 1-4-3c.

It can be seen that the higher the gain of the power-amplifier portion of the system, the better the regulation of the closed-loop system. In most systems, however, this gain cannot be increased indefinitely because of the problem of stability. This is discussed at several points in the book.

CURVE FOLLOWER. The curve follower is an example typical of many control systems. The system is shown in Fig. 1-4-4. The carriage is

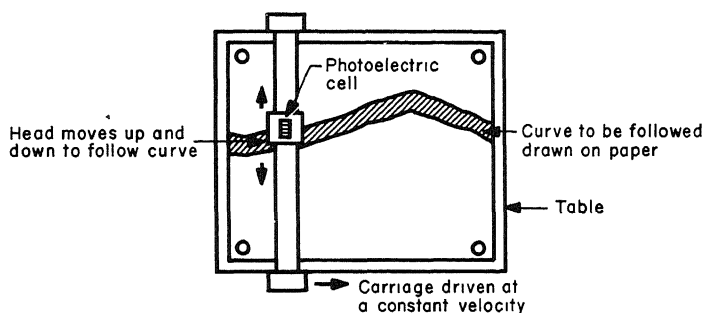


Fig. 1-4-4.

driven from left to right at a constant velocity. The curve to be followed is drawn on a piece of paper and is placed on the table top of the device. The head moves up or down on the carriage in such a way as to follow the curve drawn on the paper in the following manner. The paper is white and the curve on the paper is black. The photoelectric cell is designed so that a certain amount of white should appear below it. If more than this amount of white appears, the system responds by driving the head downward. Conversely, if less than the desired amount of white appears below it, the system responds by driving the head upward.

In the following discussion, it is assumed that the system is at rest with the carriage at the left side of the table and the head is in the desired position at the top edge of the curve. The carriage is driven at a constant velocity across the table from left to right. The head moves up or down on the carriage so that the proper amount of white appears below it. In this way the head follows the curve. If it is desired to follow a different curve, the piece of paper can be replaced with another on which the new curve is drawn.

The curve follower is ordinarily used to drive some other system. For example, the curve drawn on the paper can represent the contour (or the contour multiplied by a scale factor) to which a piece of metal is to be shaped. The output to be driven by the system may be a torch or a milling machine. In this way, the metal can be shaped to an intricate pattern by simply drawing the pattern on paper. Not only can one milling machine be driven by such a device but an entire group of machines can be made to follow the pattern.

Another modification of this system is exemplified in the ship-propeller milling machine. Instead of following a curve drawn on paper, one of the blades of the propeller is sculptured in model form. The model turns and the head mechanism moves up or down and follows the contour of the model. At the same time the propeller also turns, and its surface is milled to duplicate that of the model. After one sweep, the head and the milling machine are moved out along the radius and a second contour is milled, so that eventually the entire surface of one ship propeller blade is finished. Then the second and third blades of the propeller are milled in turn by following the same procedure, so that each blade is milled to match the surface of the same model.

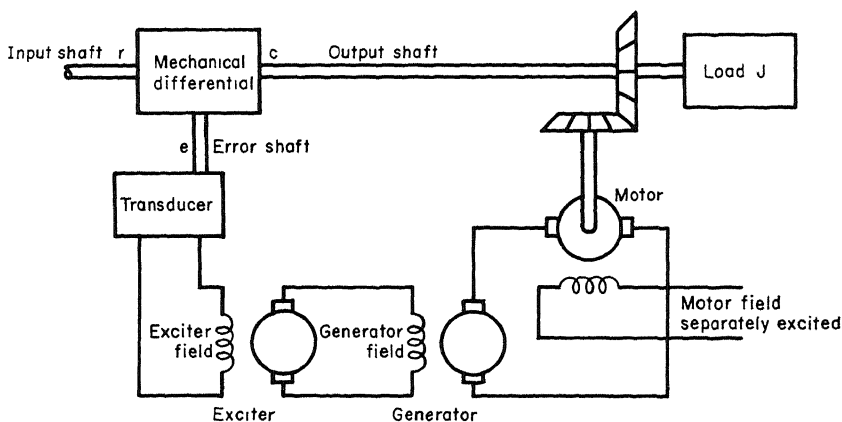


Fig. 1-4-5.

ROTATIONAL SYSTEM. The system of Fig. 1-4-5 is a more elaborate system that can be used to locate the angular position of a moment of inertia J . It is desired to have the output shaft c follow the input shaft r within a certain tolerance. The mechanical differential has three shafts. These shafts are referred to as the input shaft r , output shaft c and error shaft e . The differential is a device whereby the error shaft takes an angular position which is the difference between the position of the input and output shafts and can be described by the equation

$$e = r - c \quad (1-4-2)$$

The error shaft actuates a transducer which produces a voltage proportional to the angular displacement of the error shaft. The voltage from the transducer is applied to the field of an exciter. The output of the exciter is applied to the field of the generator. The output of the generator is applied to the motor which in turn drives the output shaft and the load.

If the system is at rest, the input and output shafts are stationary and are in alignment. From this rest position the input shaft is turned through an angle α_1 in the clockwise direction. The error shaft also turns through an angle α_1 and a voltage is produced at the output of the transducer which is applied to the field of the exciter. The exciter in turn applies a voltage to the motor, which will cause the load to turn until the output shaft is again in alignment with the input shaft.

The energy necessary to drive the load is supplied by the power devices and the energy supplied to the input is only the amount of energy necessary to turn the input shaft.

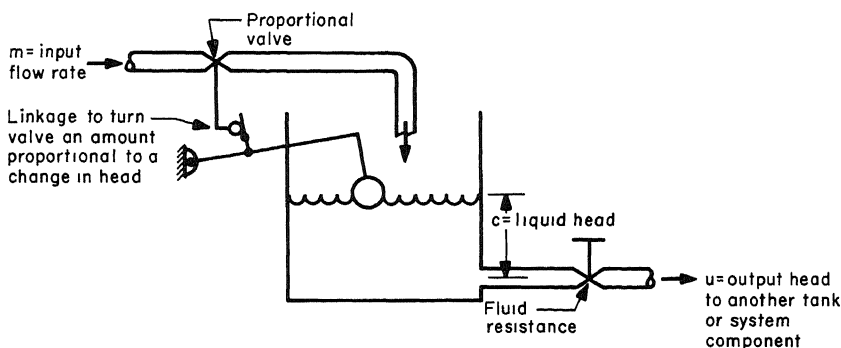


Fig. 1-4-6.

MAINTAINING LIQUID LEVEL. Fig. 1-4-6 illustrates a possible closed-loop process control system. This system is used to maintain a preset liquid level or head in a tank. The fluid level in the tank is measured by the float which is connected to the valve controlling the input flow. A drop in the head in the tank causes the linkage to open the input valve further, thereby increasing the input flow rate. This action tends to restore the original head in the tank. Therefore a closed loop exists in this system. Any disturbance that tends to cause the head in the tank to depart from a preset level initiates the closed-loop action described above. Such a disturbance might be a change from normal steady values of the output flow rate or of the input flow rate.

1-5. Linear Differential Equations. The material presented in this chapter discusses in general terms the basic concepts of control system theory. As soon as quantitative analysis is attempted, the physical laws

governing system behavior must be expressed in mathematical form. The physical laws considered in this book take the mathematical form of an ordinary differential equation of order n , where n is the order of the highest derivative that appears in the equation.

For the physical systems being considered, the desired response is usually a function of time, which is the independent variable. The dependent variable can be any physical quantity, such as voltage, current, displacement, pressure, or temperature.

An ordinary linear differential equation of order n has the form

$$a_n(t) \frac{d^n e}{dt^n} + a_{n-1}(t) \frac{d^{n-1} e}{dt^{n-1}} + \cdots + a_1(t) \frac{de}{dt} + a_0(t)e = f(t) \quad (1-5-1)$$

The equation is said to be linear because the $a_i(t)$ coefficients do not depend on e but are functions of t and the equation contains e and its derivatives as linear terms, i. e., terms of the first degree.

If $f(t)$ in Eq. 1-5-1 is set equal to zero, the resulting equation is said to be homogeneous. If $f(t)$ does not equal zero, the equation is said to be nonhomogeneous.

A solution of a differential equation is any functional relationship among the variables that satisfies the equation. This functional relationship itself must not contain any derivatives.

A particular solution of Eq. 1-5-1, is a function E_p which when substituted for e in the left side of the equation yields the particular $f(t)$ given on the right side. That is

$$a_n(t) \frac{d^n E_p}{dt^n} + a_{n-1}(t) \frac{d^{n-1} E_p}{dt^{n-1}} + \cdots + a_1(t) \frac{dE_p}{dt} + a_0(t)E_p = f(t) \quad (1-5-2)$$

A function E_1 is a solution of the homogeneous linear equation if the equation is equal to zero when E_1 is substituted for e , as shown by

$$a_n(t) \frac{d^n E_1}{dt^n} + a_{n-1}(t) \frac{d^{n-1} E_1}{dt^{n-1}} + \cdots + a_1(t) \frac{dE_1}{dt} + a_0(t)E_1 = 0 \quad (1-5-3)$$

Similarly if E_2 is a solution of the homogeneous linear equation, Eq. 1-5-3 can be rewritten with E_1 replaced by E_2 .

A more general solution of the homogeneous equation is $e = c_1 E_1 + c_2 E_2$ where c_1 and c_2 are arbitrary constants, as shown by

$$\begin{aligned} & a_n(t) \frac{d^n}{dt^n} (c_1 E_1 + c_2 E_2) + a_{n-1}(t) \frac{d^{n-1}}{dt^{n-1}} (c_1 E_1 + c_2 E_2) + \cdots \\ & + a_1(t) \frac{d}{dt} (c_1 E_1 + c_2 E_2) + a_0(t) (c_1 E_1 + c_2 E_2) \\ & = c_1 \left[a_n(t) \frac{d^n}{dt^n} (E_1) + a_{n-1}(t) \frac{d^{n-1}}{dt^{n-1}} (E_1) + \cdots + a_1(t) \frac{d}{dt} (E_1) \right. \\ & \quad \left. + a_0(t)(E_1) \right] + c_2 \left[a_n(t) \frac{d^n}{dt^n} (E_2) + a_{n-1}(t) \frac{d^{n-1}}{dt^{n-1}} (E_2) + \cdots \right. \\ & \quad \left. + a_1(t) \frac{d}{dt} (E_2) + a_0(t)(E_2) \right] = c_1[0] + c_2[0] = 0 \end{aligned} \quad (1-5-4)$$

It can be shown that the general solution of a homogeneous equation of order n is a solution containing n arbitrary constants and no solution to such an equation can contain more than n such constants. The most general solution of the homogeneous equation, called the complementary solution, contains n arbitrary constants and is given by

$$E_c = c_1 E_1 + c_2 E_2 + \cdots + c_n E_n \quad (1-5-5)$$

The general solution of the non-homogeneous equation is given by

$$e = E_c + E_p \quad (1-5-6)$$

In engineering work, the terms particular solution and complementary solution are sometimes called by other names. The particular solution is sometimes called forced response in the sense that this response is forced upon the system. The complementary solution is sometimes called the natural response in the sense that the system asserts its own natural characteristic in the solution.

In circuit theory, when the driving function is either a constant or a sinusoid, the particular solution is sometimes referred to as the steady-state component of the solution, and the complementary solution as the transient component of the solution.

Methods have been developed to obtain the general solution of a linear differential equation of the first order. Although it is impossible to find a completely general method of solving linear equations of any order, nevertheless, one can be obtained for the special case in which all the a_i terms are constants:

$$a_n \frac{d^n e}{dt^n} + a_{n-1} \frac{d^{n-1} e}{dt^{n-1}} + \cdots + a_1 \frac{de}{dt} + a_0 e = f(t) \quad (1-5-7)$$

Linear differential equations with constant coefficients are such an important class of equations, that many methods have been developed to find solutions. One method frequently used in control system literature makes use of Laplace transforms. Although by no means the only method to solve such equations, it is chosen for this book because of its wide popularity and simplicity. Chapter 2 presents some aspects of Laplace transformation theory.

1-6. Nonlinear Systems. For an air-core inductor (where the permeability is a constant independent of current), the equation that relates voltage and current is given by

$$e = L \frac{di}{dt} \quad (1-6-1)$$

This equation is a linear differential equation with a constant coefficient. The equation would still be a linear equation if L were some function of time t , the independent variable. However, if L depends on the current

i , the dependent variable, Eq. 1-6-1 is a nonlinear equation. More generally, in the physical sciences, when an equation does not have constant coefficients, the coefficients often vary as functions of the dependent variable and the resulting equation is nonlinear.

Examples of nonlinear physical elements are numerous. An iron-core inductor that possesses a B - H curve that is not a straight line has an inductance value that varies as a function of current. A resistor which has a curved volt-ampere characteristic has a resistance value that is a function of the current. A spring that normally has a restoring force directly proportional to displacement can be stretched beyond its elastic limit such that the spring becomes a nonlinear device.

As a matter of fact, every physical device is nonlinear if all possible values of the variable are considered. However, the displacement of the spring is limited so that the elastic limit is not exceeded; the spring can be considered to be a linear element. The iron-core inductor may have an air-gap of sufficient size, such that the inductor can be assumed to be linear. If a resistor is used in its normal range of power dissipation, the resistor can be assumed to be linear.

A similar reasoning process holds for most other physical devices. If the total excursion of the variable is limited to the straight-line portion of its characteristic, the device is linear.

As far as analysis and design are concerned, the advantage of being able to assume the linearity of a group of devices in a system is that the system can be described by a set of linear differential equations with constant coefficients. The method of solution of such equations is straightforward.

If a device is to operate in a system in such a way as to be a nonlinear element, the methods of manual analysis are exceedingly difficult except for the most simple cases. Modern computers have opened up new methods for analysis of nonlinear systems, with the result that it is possible to handle such systems with a reasonable effort.

It should not be inferred that just because a linear system is more easily analyzed, it is necessarily superior to a nonlinear system. However, before a person undertakes a study of nonlinear systems, he should have an excellent background in linear systems. It is for this reason, for the most part, that this book is concerned with linear systems.

PROBLEMS

1-1. For the voltage regulating system of Art. 1-4, discuss the functioning of the system in terms of the block diagram of Fig. 1-3-1b.

1-2. Devise a system similar to that shown in Fig. 1-4-3 that can be used to regulate current.

1-3. Repeat Prob. 1-1 for the system discussed in Art. 1-4.

1-4. Devise a system similar to that shown in Fig. 1-4-6 that can be used to position a mass in translatory motion.

1-5. Paper is wound in rolls in a paper mill. If the roll is turned at a constant angular velocity, the speed at which the paper moves through the mill increases as the diameter of the roll increases. Devise a closed-loop system that is designed to maintain a constant paper speed.

1-6. Classify the following differential equations:

$$(a) \quad 4 \frac{d^2 e}{dt^2} + 2 \frac{de}{dt} + 5e = 10 \sin 30t$$

$$(b) \quad 4 \frac{d^2 e}{dt^2} + 2 \frac{de}{dt} + 5e = 10$$

$$(c) \quad 4 \frac{d^2 e}{dt^2} + 2 \frac{de}{dt} + 5e = 0$$

$$(d) \quad 4t \frac{d^2 e}{dt^2} + 2t^2 \frac{de}{dt} + 5t^3 e = t^5$$

$$(e) \quad 4t \frac{d^2 e}{dt^2} + 2t^2 \frac{de}{dt} + 5t^3 e^2 = t^5$$

1-7. Devise a simple closed-loop system that might be used in a mechanical manipulator for remotely handling radioactive materials.

1-8. For a system described by the differential equation

$$\frac{d^2 x}{dt^2} + 25 \frac{dx}{dt} + 100 x = f(t)$$

show that if $f(t) = f_1(t) = 1.0$ for $t \geq 0$, then

$$x_1(t) = 0.01 - \frac{1}{15} e^{-5t} + \frac{1}{300} e^{-20t}$$

is the solution. Also show that if $f(t) = f_2(t) = \sin 10t$, then

$$x_2(t) = \frac{2}{475} e^{-5t} - \frac{1}{750} e^{-20t} - 0.004 \cos 10t$$

is the solution. If $f(t) = f_1(t) + f_2(t)$, is $x_1(t) + x_2(t)$ the solution of the above differential equation?

2

THE LAPLACE TRANSFORMATION

2-1. Introduction. This chapter presents some of the basic principles of Laplace transformation theory. Since the remainder of the book is concerned with applications, they are not featured in this chapter.

Laplace transformation methods offer a systematic procedure whereby a linear differential equation with constant coefficients and its associated initial conditions can be solved in a straightforward manner.

2-2. The Direct Transform Integral. By definition, the Laplace transform of a function of time $f(t)$ is obtained by multiplying $f(t)$ by ε^{-st} and integrating the product between the limits zero and infinity:

$$\mathcal{L}[f(t)] = \int_0^{\infty} f(t)\varepsilon^{-st} dt = F(s) \quad (2-2-1)$$

The symbol $\mathcal{L}[f(t)]$ is read as the Laplace transform of $f(t)$. As far as the integration is concerned, s is a complex constant. Because of the limits, t does not appear in the result. Therefore, the result is a function of s and is denoted as $F(s)$.

Since Eq. 2-2-1 is an improper integral, the integral does not exist unless certain conditions are imposed on $f(t)$. One permissible set of conditions is that $f(t)$ be of exponential order and be sectionally continuous over the range $0 < t < \infty$. Fortunately, these conditions are usually satisfied by functions encountered in the physical sciences.

A function $f(t)$ is of exponential order if a real positive constant σ exists such that the product

$$\varepsilon^{-\sigma t}|f(t)| \quad (2-2-2)$$

has the limit zero as t approaches infinity. For each $f(t)$ this requirement places a limitation on the possible values of σ . The value σ_0 , termed the abscissa of convergence, is defined such that this limit is zero if σ is greater than σ_0 . For example, if $f(t) = \varepsilon^{+2t}$, then $\lim \varepsilon^{-\sigma t}|\varepsilon^{+2t}|$ approaches zero only if σ is greater than 2. Therefore, σ_0 for this $f(t)$ is equal to 2.

An example of an $f(t)$ that is not of exponential order is

$$f(t) = \varepsilon^{+t^2} \quad (2-2-3)$$

Therefore, the integral in Eq. 2-2-1 does not exist for this $f(t)$ and hence this $f(t)$ does not have a Laplace transform.

If the integral exists for an $f(t)$, the Laplace method transforms $f(t)$ into an $F(s)$ which is a function of the complex variable s . The complex variable has a real part σ and an imaginary part $j\omega$, so that s is given by

$$s = \sigma + j\omega \quad (2-2-4)$$

The complex s -plane has σ represented on the axis of the reals and $j\omega$ along the imaginary axis. Frequencies (as the word is conventionally interpreted) produced by an oscillator appear on the $j\omega$ -axis. The use of the complex s -plane is carried throughout the book and is explained in detail in Art. 2-1.

The direct Laplace transform is said to transform from the time domain $f(t)$ to the frequency domain $F(s)$. The inverse Laplace transformation is indicated by

$$\mathcal{L}^{-1}[F(s)] = f(t) \quad (2-2-5)$$

where the symbol $\mathcal{L}^{-1}[F(s)]$ is read as the inverse Laplace transform of $F(s)$. The inverse transform is said to transform from the frequency domain $F(s)$ to the time domain $f(t)$. More is said about the inverse transform in Art. 2-8.

2-3. Transform Theorems. The following theorems are basic in the theory of the Laplace transformation.

THEOREM 2-3-1. The transform of the product of a constant and a function is the product of the constant and the transform of this function. This is shown as

$$\mathcal{L}[Kf(t)] = K\mathcal{L}[f(t)] \quad (2-3-1)$$

Since the direct transform is defined as an integral and the constant K can be moved outside the integral sign, the proof of this theorem is evident.

THEOREM 2-3-2. The transform of a sum (or difference) of time functions is the sum (or difference) of the transforms of the individual time functions when these latter exist. This is shown as

$$\mathcal{L}[f_1(t) + f_2(t) - f_3(t)] = \mathcal{L}[f_1(t)] + \mathcal{L}[f_2(t)] - \mathcal{L}[f_3(t)] \quad (2-3-2)$$

This relationship is again true because the direct transform is defined as an integral and the integral of a sum is the sum of the integrals, provided that each of the individual integrals exists.

THEOREM 2-3-3. The transform of the derivative of a function is given formally by

$$\mathcal{L}[f'(t)] = \int_0^{\infty} f'(t)e^{-st} dt \quad (2-3-3)$$

which can be evaluated by using integration by parts

$$\int_0^\infty u \, dv = uv \Big|_0^\infty - \int_0^\infty v \, du \quad (2-3-4)$$

with the following identifications:

$$\begin{aligned} u &= \varepsilon^{-st} & dv &= f'(t) \, dt \\ du &= -s\varepsilon^{-st} \, dt & v &= f(t) \end{aligned} \quad (2-3-5)$$

Upon substitution, the integral becomes

$$\mathcal{L}[f'(t)] = [f(t)\varepsilon^{-st}] \Big|_0^\infty - \int_0^\infty [-s\varepsilon^{-st}]f(t) \, dt \quad (2-3-6)$$

which can be written as

$$\mathcal{L}[f'(t)] = \lim_{t \rightarrow \infty} [f(t)\varepsilon^{-st}] - f(0)\varepsilon^{-0} + s \int_0^\infty f(t)\varepsilon^{-st} \, dt \quad (2-3-7)$$

If $f(t)$ is of exponential order and σ (the real part of s) is greater than σ_0 , then the term

$$\lim_{t \rightarrow \infty} [f(t)\varepsilon^{-st}] \quad (2-3-8)$$

approaches zero. Therefore, Eq. 2-3-7 can be written as

$$\mathcal{L}[f'(t)] = s \mathcal{L}[f(t)] - f(0) \quad (2-3-9)$$

where $f(0)$ is the limit of $f(t)$ as t approaches zero through positive values. In many places in the literature, $f(0)$ is written as $f(0^+)$ and $t = 0$ as $t = 0^+$ to emphasize that both limits are taken as t approaches zero through positive values. The $+$ sign is implied but not written in this book.

THEOREM 2-3-4. The transform of the second derivative $\mathcal{L}[f''(t)]$ can be found by using Eq. 2-3-9 and the following technique.

The function $g(t)$ is defined as

$$g(t) = f'(t) \quad (2-3-10)$$

When $g(t)$ is differentiated, the result is

$$g'(t) = f''(t)$$

Therefore, the Laplace transform of the second derivative in terms of $g(t)$ is

$$\mathcal{L}[f''(t)] = \mathcal{L}[g'(t)] = s \mathcal{L}[g(t)] - g(0) \quad (2-3-11)$$

In terms of $f(t)$ and its derivative, the equation is

$$\begin{aligned} \mathcal{L}[f''(t)] &= s \{ \mathcal{L}[f'(t)] \} - f'(0) \\ &= s \{ s \mathcal{L}[f(t)] - f(0) \} - f'(0) \\ &= s^2 \mathcal{L}[f(t)] - sf(0) - f'(0) \end{aligned} \quad (2-3-12)$$

By extending this technique, the transform for the third-order derivative can be found as

$$\mathcal{L}[f'''(t)] = s^3 \mathcal{L}[f(t)] - s^2 f(0) - s f'(0) - f''(0) \quad (2-3-13)$$

By use of mathematical induction, the form for the transform of a derivative of any order can be seen to be

$$\mathcal{L}[f^{(n)}(t)] = s^n \mathcal{L}[f(t)] - s^{n-1} f(0) - s^{n-2} f'(0) - \cdots - f^{(n-1)}(0) \quad (2-3-14)$$

For the transform of each of these higher-order derivatives to exist, $f^{(n)}(t)$ must be such that it has a Laplace transform. That is, the $f^{(n)}(t)$ is sectionally continuous and is of exponential order.

THEOREM 2-3-5. The transform of the integral of $f(t)$ between 0 and t can be found by using integration by parts on the direct transform integral

$$\mathcal{L}[f(t)] = \int_0^\infty f(t) \varepsilon^{-st} dt \quad (2-3-15)$$

by making the following identifications:

$$\begin{aligned} dv &= f(t) dt & u &= \varepsilon^{-st} \\ v &= \int_0^t f(t) dt & du &= -s \varepsilon^{-st} dt \end{aligned}$$

Upon substitution, the form

$$\int_0^\infty u dv = uv \Big|_0^\infty - \int_0^\infty v du$$

becomes

$$\mathcal{L}[f(t)] = \left[\varepsilon^{-st} \int_0^t f(t) dt \right]_0^\infty - \int_0^\infty \left[\int_0^t f(t) dt \right] [-s \varepsilon^{-st} dt] \quad (2-3-16)$$

which can be written as

$$\mathcal{L}[f(t)] = \lim_{t \rightarrow \infty} \varepsilon^{-st} \int_0^t f(t) dt - \varepsilon^{-s0} \int_0^0 f(t) dt + s \int_0^\infty \left[\int_0^t f(t) dt \right] \varepsilon^{-st} dt \quad (2-3-17)$$

or

$$\mathcal{L}[f(t)] = s \mathcal{L} \left[\int_0^t f(t) dt \right] \quad (2-3-18)$$

From this the transform of the definite integral is found to be

$$\mathcal{L} \left[\int_0^t f(t) dt \right] = \frac{\mathcal{L}[f(t)]}{s} \quad (2-3-19)$$

In obtaining Eq. 2-3-19, it is again assumed that $f(t)$ is of exponential order so that the term

$$\lim_{t \rightarrow \infty} \varepsilon^{-st} \int_0^t f(t) dt \quad (2-3-20)$$

equals zero. If $f(t)$ is of exponential order, the integral of $f(t)$ is also of exponential order and as t approaches infinity, the term in Eq. 2-3-20 approaches zero provided the real part of s is greater than σ_0 .

THEOREM 2-3-6. The transform of a multiple integral can be determined from Eq. 2-3-19 in much the same that the transform of a multiple derivative is determined from Eq. 2-3-9.

The double integral

$$\int_0^t \int_0^t f(t) dt dt \quad (2-3-21)$$

can be written as

$$\int_0^t h(t) dt \quad (2-3-22)$$

where

$$h(t) = \int_0^t f(t) dt \quad (2-3-23)$$

The transform for the double integral is

$$\begin{aligned} \mathcal{S}\left[\int_0^t \int_0^t f(t) dt dt\right] &= \mathcal{S}\left[\int_0^t h(t) dt\right] \\ &= \frac{\mathcal{S}[h(t)]}{s} = \frac{\mathcal{S}\left[\int_0^t f(t) dt\right]}{s} = \frac{\mathcal{S}[f(t)]}{s^2} \end{aligned} \quad (2-3-24)$$

This process can be continued until the general term is obtained

$$\mathcal{S}\left[\int_0^t \int_0^t \cdots \int_0^t f(t) dt^n\right] = \frac{\mathcal{S}[f(t)]}{s^n} \quad (2-3-25)$$

THEOREM 2-3-7. The transform of $\varepsilon^{-at}f(t)$ for an arbitrary a can be determined from the transform of $f(t)$, assuming that $f(t)$ has a transform. The transform of $f(t)$ is indicated by

$$\int_0^\infty f(t)\varepsilon^{-st} dt = F(s) \quad (2-3-26)$$

The transform of $\varepsilon^{-at}f(t)$ then is

$$\begin{aligned} \mathcal{S}[\varepsilon^{-at}f(t)] &= \int_0^\infty \varepsilon^{-at}f(t)\varepsilon^{-st} dt \\ &= \int_0^\infty f(t)\varepsilon^{-(s+a)t} dt = F(s+a) \end{aligned} \quad (2-3-27)$$

Eqs. 2-3-26 and 2-3-27 can be interpreted in the following manner. If $f(t)$ has a transform $F(s)$, then $\varepsilon^{-at}f(t)$ has a transform which can be obtained from $F(s)$ by replacing s by $s + a$.

THEOREM 2-3-8. If $F(s)$ multiplied by ε^{-Ts} , the time function $f_T(t)$ which has the transform $\varepsilon^{-Ts}F(s)$ can be determined from the original $f(t)$ by translating the graph of $f(t)$ to the right through a distance of T units with $f_T(t)$ equal to zero between $t = 0$ and $t = T$. This type of translation is sometimes referred to as a dead-time delay. The significance of this is shown in the following manner. The original $f(t)$ and $F(s)$ are related by the direct transform integral

$$F(s) = \int_0^{\infty} f(t)\varepsilon^{-st} dt \quad (2-3-28)$$

Both sides of this equation can be multiplied by ε^{-Ts}

$$\varepsilon^{-Ts}F(s) = \varepsilon^{-Ts} \int_0^{\infty} f(t)\varepsilon^{-st} dt \quad (2-3-29)$$

As far as the integral portion of Eq. 2-3-29 is concerned, ε^{-Ts} is a constant and can be moved inside the integral sign as

$$\varepsilon^{-Ts}F(s) = \int_0^{\infty} f(t)\varepsilon^{-s(t+T)} dt \quad (2-3-30)$$

This equation can be rewritten with a change of variable

$$\tau = t + T \quad (2-3-31)$$

The new limits of integration and the differential are

$$\begin{aligned} t = 0 & ; \tau = T \\ t = \infty & ; \tau = \infty \\ dt & = d\tau \end{aligned} \quad (2-3-32)$$

The equation with the change of variable is

$$\varepsilon^{-Ts}F(s) = \int_0^T (0) d\tau + \int_T^{\infty} f(\tau - T)\varepsilon^{-s\tau} d\tau \quad (2-3-33)$$

The new function $f_T(t)$ can be determined from this equation as

$$\begin{aligned} f_T(t) &= 0; \quad 0 < t < T \\ f_T(t) &= f(t - T); \quad T < t < \infty \end{aligned} \quad (2-3-34)$$

Hence the theorem is demonstrated that if $\mathcal{L}[f(t) u_{-1}(t)] = F(s)$ then $\mathcal{L}[f(t - T)u_{-1}(t - T)] = \varepsilon^{-Ts}F(s)$, where $u_{-1}(t - T)$ is the unit-step function occurring at $t = T$. In the next article, a few examples demonstrate the usefulness of this theorem.

THEOREM 2-3-9. The transform of a periodic wave shape can be obtained by the following procedure. The wave shape is periodic and the transform of the function $f_1(t)$ which has the shape of the original wave in the first period and is zero thereafter can be obtained and is denoted $F_1(s)$. Since the wave shape in the second period is exactly the same as that in the first, the wave shape in the second period can be thought of as being the wave shape in the first period delayed by one period and therefore has the transform $F_1(s)\varepsilon^{-Ts}$ where T is the time of the period of the repeating wave shape. By a similar reasoning process, the transform of the wave shape in the third period is $F_1(s)\varepsilon^{-2Ts}$. By extending this method, the transform for the periodic wave shape that is denoted by $F(s)$ can be found as

$$F(s) = F_1(s)[1 + \varepsilon^{-Ts} + \varepsilon^{-2Ts} + \varepsilon^{-3Ts} + \dots] \quad (2-3-35)$$

The infinite series inside the brackets can be summed for $|\varepsilon^{-Ts}| < 1$ so that $F(s)$ is given by

$$F(s) = \frac{F_1(s)}{1 - \varepsilon^{-Ts}} \text{ for } |\varepsilon^{-Ts}| < 1 \quad (2-3-36)$$

This technique for obtaining the transform of a periodic function can be stated as follows: The repeating wave shape is examined and the period T of the wave is determined; the transform $F_1(s)$ of the function $f_1(t)$ which is the form of the first period of the wave shape and is zero thereafter is then found; $F(s)$ the transform of the repeating wave is obtained by multiplying $F_1(s)$ by the factor $1/(1 - \varepsilon^{-Ts})$. In the next article a few examples demonstrate the use of this technique.

THEOREM 2-3-10. The multiplication of two transforms such as

$$F_1(s)F_2(s) \quad (2-3-37)$$

gives the transform of a function which results from what is called the convolution of the two corresponding time functions. $F_1(s)$ and $F_2(s)$ are the transforms of $f_1(t)$ and $f_2(t)$ respectively as given by

$$\begin{aligned} \mathcal{L}[f_1(t)] &= F_1(s) \\ \mathcal{L}[f_2(t)] &= F_2(s) \end{aligned} \quad (2-3-38)$$

The time function which has the transform $F_1(s)F_2(s)$ is the convolution of $f_1(t)$ and $f_2(t)$ as given by

$$\mathcal{L}^{-1}[F_1(s)F_2(s)] = \int_0^t f_1(t - \tau)f_2(\tau) d\tau = f_1(t)*f_2(t) \quad (2-3-39)$$

or by

$$\mathcal{L}^{-1}[F_2(s)F_1(s)] = \int_0^t f_2(t - \tau)f_1(\tau) d\tau = f_2(t)*f_1(t) \quad (2-3-40)$$

The symbols $f_1(t)*f_2(t)$ and $f_2(t)*f_1(t)$ denote the two forms of the convolution integral. The function with the argument $t - \tau$ in the convolution integral is written first. Examples of the use of the convolution integral and a geometric interpretation of the integral are presented in Chapter 10.

THEOREM 2-3-11. An $f(t)$ is given which has a Laplace transform $F(s)$:

$$\mathcal{L}[f(t)] = \int_0^{\infty} f(t)\varepsilon^{-st} dt = F(s) \quad (2-3-41)$$

The t in $f(t)$ is replaced by a linear transformation at ($a > 0$), and the effect upon $F(s)$ is to be determined.

$$\mathcal{L}[f(at)] = \int_0^{\infty} f(at)\varepsilon^{-st} dt \quad (2-3-42)$$

With a change of variable

$$at = \tau \quad (2-3-43)$$

the equation becomes

$$\mathcal{L}[f(at)] = \frac{1}{a} \int_0^{\infty} f(\tau)\varepsilon^{-s\tau/a} d\tau = \frac{1}{a} F \frac{s}{a} \quad (2-3-44)$$

This result is the fundamental theorem upon which scaling is based. The subject of scaling is discussed at several places in the book.

THEOREM 2-3-12. Differentiation of the transform of a function with respect to s corresponds to the multiplication of the time function by $-t$. The $f(t)$ and the corresponding $F(s)$ are given as

$$F(s) = \int_0^{\infty} \varepsilon^{-st} f(t) dt \quad (2-3-45)$$

The derivative of $F(s)$ is given by

$$\frac{dF(s)}{ds} = \frac{\partial}{\partial s} \int_0^{\infty} \varepsilon^{-st} f(t) dt \quad (2-3-46)$$

If $f(t)$ is sectionally continuous and of exponential order, the differentiation can be performed under the integral sign as

$$\frac{dF(s)}{ds} = \int_0^{\infty} \frac{\partial}{\partial s} [\varepsilon^{-st} f(t)] dt = \int_0^{\infty} \varepsilon^{-st} [-tf(t)] dt \quad (2-3-47)$$

which can be written as

$$\frac{dF(s)}{ds} = \mathcal{L}[-tf(t)] \quad (2-3-48)$$

This process can be generalized as

$$\frac{d^n F(s)}{ds^n} = \mathcal{L}[(-t)^n f(t)] \quad (\text{where } n = 1, 2, \dots) \quad (2-3-49)$$

2-4. Development of Transform Pairs. The following examples illustrate the theorems of the previous article.

Example 2-4-1. The unit-step function $u_{-1}(t)$ is defined by

$$\begin{aligned} u_{-1}(t) &= 0, t < 0 \\ u_{-1}(t) &= 1, t > 0 \end{aligned} \quad (2-4-1)$$

Where circuit analysis is concerned, this function represents the closing of a switch at $t = 0$ on a battery of unit voltage or the opening of a switch at $t = 0$ on a current source of unit current. Similarly, in mechanical systems, the unit-step function represents the application to the system of a unit force or of a unit torque at $t = 0$. Where control systems are concerned, the unit-step function often represents the application of a unit displacement to the input of the system at $t = 0$. Many other such examples could be used.

Since in the direct transform integral

$$\int_0^{\infty} f(t) e^{-st} dt$$

the values of $f(t)$ for negative time are immaterial, the $u_{-1}(t)$ has the value 1 as far as the integral is concerned and the direct transform of this function is found as indicated:

$$F(s) = \int_0^{\infty} 1 e^{-st} dt = -\frac{1}{s} e^{-st} \Big|_0^{\infty} = -\frac{1}{s} \lim_{t \rightarrow \infty} e^{-st} - e^{-s \cdot 0} = \frac{1}{s} \quad (2-4-2)$$

In performing the integration, the term $\lim_{t \rightarrow \infty} e^{-st}$ is assumed to be zero. This will be true only if the real part of s is greater than zero. To examine this in some detail, it is assumed that $s = -2$. The integral

$$\int_0^{\infty} e^{+2t} dt$$

does not exist. However, if $s = +2$, the integral

$$\int_0^{\infty} e^{-2t} dt$$

does exist. Instead of letting $s = +2$, it can be set equal to σ (the real part of s) and the integral

$$\int_0^{\infty} e^{-\sigma t} dt$$

converges as long as σ is greater than zero. This concept can be broadened by letting s take on complex values $s = \sigma + j\omega$; then the integral

$$\int_0^{\infty} e^{-st} dt = \frac{1}{s} = F(s) \quad (2-4-3)$$

exists in the half-plane to the right of $\sigma = 0$.

In general the abscissa of convergence is thus determined and is denoted by σ_0 . In this particular example $\sigma_0 = 0$.

Example 2-4-2. The transform of $\varepsilon^{-at}u_{-1}(t)$ is considered next. When any function such as ε^{-at} is multiplied by $u_{-1}(t)$, the resulting function is zero for negative time. If a system is energized at $t = 0$ by a driving function, all response functions should be multiplied by $u_{-1}(t)$ if the expression is to hold for all values of t . For example, if a response function to an input applied at $t = 0$ is of the form

$$\varepsilon^{-at} \sin(\omega t + \delta)$$

the expression should be written as

$$[\varepsilon^{-at} \sin(\omega t + \delta)]u_{-1}(t)$$

if the response is to be expressed correctly for $t < 0$. However, the factor $u_{-1}(t)$ usually is understood and not written explicitly

The transform of ε^{-at} can be found by use of the direct integral

$$\int_0^{\infty} \varepsilon^{-at} \varepsilon^{-st} dt = F(s) \quad (2-4-4)$$

or by use of the theorem developed in Art. 2-3-6

$$\mathcal{L}[\varepsilon^{-at}f(t)] = F(s + a) \quad (2-4-5)$$

If the second of these methods is used, the $f(t)$ can be identified as 1, which has a transform $1/s$. The transform of ε^{-at} can be found from $1/s$ by replacing s by $s + a$. Therefore

$$\mathcal{L}[\varepsilon^{-at}] = \frac{1}{s + a} \quad (2-4-6)$$

The abscissa of convergence σ_0 can be determined from the $F(s)$ of Eq. 2-4-6, as is done in Art. 2-7. However, if s is allowed to take on its real part σ , Eq. 2-4-4 becomes

$$\int_0^{\infty} \varepsilon^{-(\sigma + a)t} dt$$

By inspection, it can be seen that this integral will converge if σ is greater than $-a$. Therefore it has been determined that $\sigma_0 = -a$.

Example 2-4-3. The transform of $\sin \omega t$ can be found by use of the defining integral

$$\int_0^{\infty} \sin \omega t \varepsilon^{-st} dt = F(s) \quad (2-4-7)$$

By expanding $\sin \omega t$ in its exponential form as

$$\sin \omega t = \frac{\varepsilon^{+j\omega t} - \varepsilon^{-j\omega t}}{2j} \quad (2-4-8)$$

the transform is found as

$$\begin{aligned} \mathcal{L}[\sin \omega t] &= \mathcal{L}\left[\frac{\varepsilon^{+j\omega t} - \varepsilon^{-j\omega t}}{2j}\right] \\ &= \frac{1}{2j} \left\{ \mathcal{L}[\varepsilon^{+j\omega t}] - \mathcal{L}[\varepsilon^{-j\omega t}] \right\} \end{aligned}$$

Eq. 2-4-6 can be used, with $a = -j\omega$ for the first term and $a = +j\omega$ for the second term, giving

$$\mathcal{L}[\sin \omega t] = \frac{1}{2j} \left\{ \frac{1}{s - j\omega} - \frac{1}{s + j\omega} \right\} = \frac{\omega}{s^2 + \omega^2} \quad (2-4-9)$$

By inspection, it is evident that $\sigma_0 = 0$.

Example 2-4-4. The transform of $\varepsilon^{-at} \sin \omega t$ is found from $\mathcal{L}[\sin \omega t]$ by replacing s by $s + a$. Therefore,

$$\mathcal{L}[\varepsilon^{-at} \sin \omega t] = \frac{\omega}{(s + a)^2 + \omega^2} \quad (2-4-10)$$

The value for σ_0 can be shown to be $\sigma_0 = -a$.

Example 2-4-5. The transform of $\cos \omega t$ is found by starting from

$$\cos \omega t = \frac{1}{\omega} \frac{d}{dt}(\sin \omega t)$$

and taking the transform of both sides as

$$\mathcal{L}[\cos \omega t] = \frac{1}{\omega} [s \mathcal{L}(\sin \omega t) - 0]$$

which leads to

$$\mathcal{L}[\cos \omega t] = \frac{s}{s^2 + \omega^2} \quad (2-4-11)$$

For this transform, $\sigma_0 = 0$.

Example 2-4-6. The transform of $\varepsilon^{-at} \cos \omega t$ can be found from $\mathcal{L}[\cos \omega t]$ by replacing s by $s + a$. Therefore,

$$\mathcal{L}[\varepsilon^{-at} \cos \omega t] = \frac{s + a}{(s + a)^2 + \omega^2} \quad (2-4-12)$$

For this transform, $\sigma_0 = -a$.

Many similar examples could be included at this point. However, the few examples presented demonstrate several of the methods by which the direct transformation can be made. An extensive table of transform pairs is included in the Appendix.

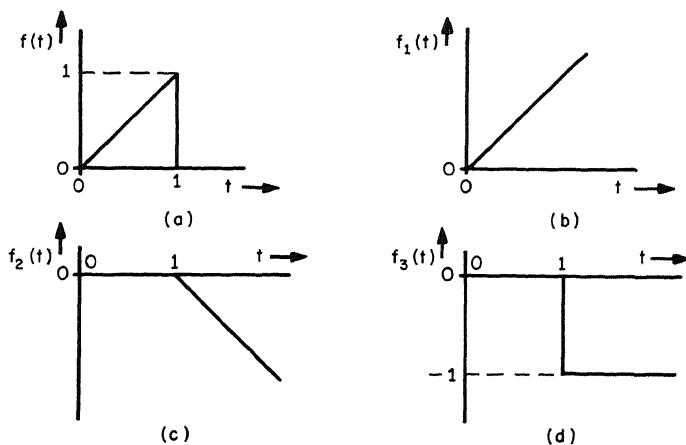


Fig. 2-5-1.

2-5. Use of the Dead-Time Delay Operator. The dead-time delay operator can sometimes be employed in obtaining transforms.

Example 2-5-1. Certain functions can be made up from simple functions. An example of this is the function shown in Fig. 2-5-1a. This function can be obtained by adding the functions of Fig. 2-5-1b, c, and d. The transform of the function shown in Fig. 2-5-1b is

$$\frac{1}{s^2} \quad (2-5-1)$$

If the function shown in Fig. 2-5-1c started at the origin, its transform would be $-1/s^2$. However, since it starts at 1 instead of at the origin, this $F(s)$ must be multiplied by $e^{-T}s$ which is the operator for a dead-time delay of T seconds. In this case $T = 1$; therefore, the transform for the function in Fig. 2-5-1c is

$$-\frac{1}{s^2}e^{-s} \quad (2-5-2)$$

If a similar line of reasoning is followed, the transform for the function in Fig. 2-5-1d is

$$-\frac{1}{s}e^{-s} \quad (2-5-3)$$

The transform for the original function in Fig. 2-5-1a is the sum of the three transforms as given by

$$F(s) = \frac{1}{s^2} - \frac{1}{s^2}e^{-s} - \frac{1}{s}e^{-s} \quad (2-5-4)$$

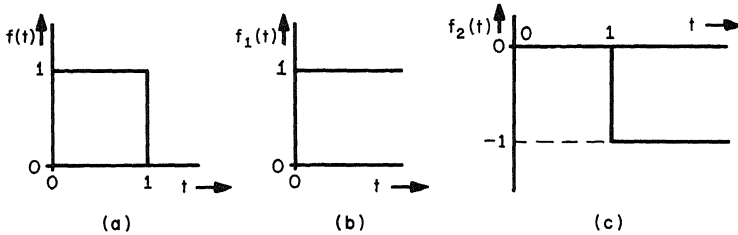


Fig. 2-5-2.

Example 2-5-2. The next example is the rectangular pulse shown in Fig. 2-5-2a. This function can be formed by adding the two components of Fig. 2-5-2b and c. The transform of the rectangular pulse of Fig. 2-5-2a therefore becomes

$$F(s) = \frac{1}{s} - \frac{1}{s}e^{-s} \quad (2-5-5)$$

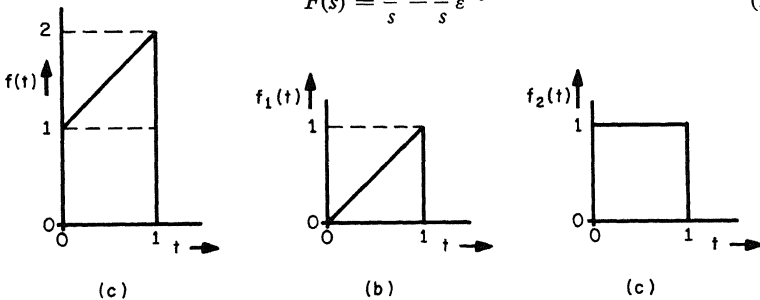


Fig. 2-5-3.

Example 2-5-3. The wave form shown in Fig. 2-5-3a is considered next. This function is formed by adding the two wave shapes of Fig. 2-5-3b and c of which the transforms are determined in the last two examples. Therefore, the transform of the wave shape of Fig. 2-5-3a is the sum of Eqs. 2-5-4 and 2-5-5:

$$F(s) = \frac{1}{s^2} + \frac{1}{s} - \frac{1}{s^2} \varepsilon^{-s} - \frac{2}{s} \varepsilon^{-s} \quad (2-5-6)$$

Example 2-5-4. The transform of a periodic wave shape can be determined by use of Theorem 2-3-8. This concept is reviewed briefly: the periodic function is examined and the period T of the wave shape is determined; the transform of the wave shape for the first period is determined as $F_1(s)$; the transform $F(s)$ of the periodic wave shape is obtained by multiplying $F_1(s)$ by the factor $1/(1 - \varepsilon^{-Ts})$. The wave

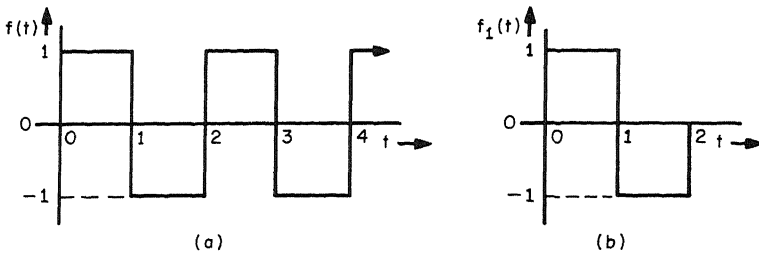


Fig. 2-5-4.

shape of Fig. 2-5-4a is used as an example. The period T can be seen to be 2 seconds and Fig. 2-5-4b shows $f_1(t)$, which includes the first period of the wave form. The three components that make up $f_1(t)$ are readily visualized and $F_1(s)$ determined as

$$F_1(s) = \frac{1}{s} - \frac{2}{s} \varepsilon^{-s} + \frac{1}{s} \varepsilon^{-2s} = \frac{(1 - \varepsilon^{-s})^2}{s} \quad (2-5-7)$$

The $F_1(s)$ is multiplied by $1/(1 - \varepsilon^{-Ts})$ to obtain $F(s)$ as

$$\begin{aligned} F(s) &= \frac{(1 - \varepsilon^{-s})^2}{s(1 - \varepsilon^{-2s})} = \frac{(1 - \varepsilon^{-s})^2}{s(1 - \varepsilon^{-s})(1 + \varepsilon^{-s})} \\ &= \frac{1 - \varepsilon^{-s}}{s(1 + \varepsilon^{-s})} = \frac{\varepsilon^{+s/2} - \varepsilon^{-s/2}}{s(\varepsilon^{+s/2} + \varepsilon^{-s/2})} = \frac{1}{s} \tanh \frac{s}{2} \end{aligned} \quad (2-5-8)$$

Example 2-5-5. The transform of a unit impulse is found after the concept of the unit impulse is reviewed. The latter can be introduced in many ways. For the present purpose, the rectangular pulse shown in Fig. 2-5-5a is the starting point.

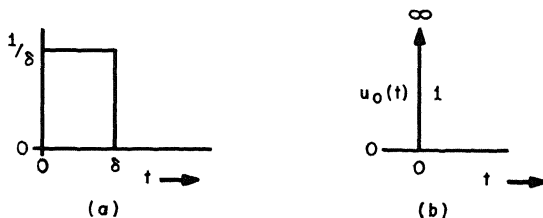


Fig. 2-5-5.

It can be seen that the area under the curve of the pulse is equal to unity regardless of the value of δ . As δ approaches zero as a limit, the height of the pulse increases without bound. In the limit the pulse is an impulse of unit area. Hence the term unit-impulse. The symbol for the unit impulse is $u_0(t)$. The pictorial representation of the unit impulse is shown in Fig. 2-5-5b. The ∞ symbol above the arrow and the 1 beside the arrow indicate, respectively, that the magnitude of the function is infinite and that its area is unity.

The transform of the rectangular pulse of Fig. 2-5-5a is

$$\frac{1}{\delta} \left[\frac{1 - e^{-\delta s}}{s} \right] \quad (2-5-9)$$

As δ approaches zero, Eq. 2-5-9 approaches the transform of the unit impulse. If the limit is taken as the equation stands, the result is the indeterminate form $0/0$ which, by l'Hospital's rule, can be evaluated by differentiating the numerator and the denominator with respect to δ and then taking the limit.

$$\mathcal{L}[u_0(t)] = \lim_{\delta \rightarrow 0} \frac{1}{\delta} \left[\frac{1 - e^{-\delta s}}{s} \right] = \lim_{\delta \rightarrow 0} \frac{s e^{-\delta s}}{s} = 1 \quad (2-5-10)$$

It should be added immediately that the rectangular pulse of Fig. 2-5-5a is not the only geometric form from which the unit-impulse function can be derived. Actually, any wave shape could be used, subject to the limitation that the area under the curve be unity throughout the process of taking the limit as the width δ of the pulse approaches zero. Each of

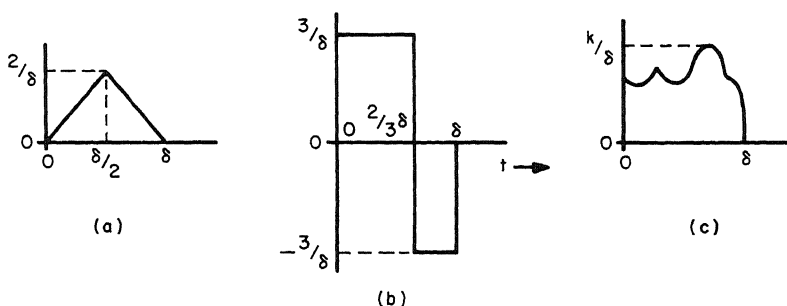


Fig. 2-5-6.

the three functions shown in Fig. 2-5-6 along with an infinite number of other functions can be used in obtaining the impulse function.

The impulse function has many uses in analysis and design. The impulse response of a physical system is the basis of several analytical procedures which are discussed at the appropriate points throughout the book. The impulse function is convenient also for introducing certain types of initial conditions into a physical system. This concept also is examined at several points in the book.

The one aspect of the impulse function to be considered at this point is the manner in which this function can extend the concept of differentia-

tion. At a point of discontinuity of a function, it has been customary to speak of the left-hand derivative and the right-hand derivative but to say that at the point of discontinuity the derivative is not defined. The following discussion indicates how the impulse broadens this concept.

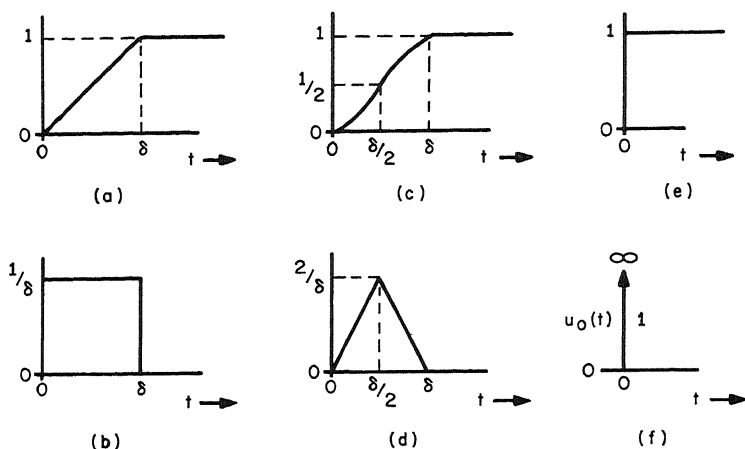


Fig. 2-5-7.

When the function shown in Fig. 2-5-7a is differentiated, the result is that shown in Fig. 2-5-7b. If δ is allowed to approach zero, the curve of Fig. 2-5-7a approaches the curve of Fig. 2-5-7e and has a unit discontinuity at $t = 0$.

The curve of Fig. 2-5-7b is a rectangular pulse and as δ approaches zero this curve approaches the unit impulse shown in Fig. 2-5-7f. Therefore the unit impulse can be thought of as representing the derivative at the point of discontinuity of the function with a unit discontinuity.

This same line of reasoning can be repeated by starting with the function of Fig. 2-5-7c which has the derivative shown in Fig. 2-5-7d. As δ approaches zero, the curve of Fig. 2-5-7c approaches that of Fig. 2-5-7e while the curve of Fig. 2-5-7d approaches the curve of Fig. 2-5-7f. The zero value at $t = 0$ and the unit value at $t = \delta$ for the curve of Fig. 2-5-7c could be connected with any other function, and its derivative in the limit would again yield the unit impulse at the point of discontinuity.

Before leaving this matter, it is of interest to think of the curve of Fig. 2-5-7e as being the integral of the curve of Fig. 2-5-7f. Integration of a curve can be visualized as the area under the curve to the left of the time t as t moves from left to right across the curve. The area under an impulse curve is zero until the impulse function occurs at $t = 0$, and thereafter the area is unity. Therefore the integration of the unit impulse of Fig. 2-5-7f leads to the unit-step function of Fig. 2-5-7e.

The unit impulse adds to the completeness of differentiation in that a function containing a number of discontinuities can be differentiated. At

each discontinuity an impulse is generated. The value of each impulse is numerically equal to the magnitude of each respective discontinuity. Since attention is being centered on differentiation, the question that arises is the effect of differentiation upon an impulse. This is explored in the next example.

Example 2-5-6. Before finding the transform of the derivative of the unit impulse, the concept is reviewed.

Since the unit impulse can be obtained from an infinite number of wave shapes, it seems reasonable to choose the wave shape which lends itself most conveniently to differentiation. The wave shape chosen here is shown in Fig. 2-5-8a, and its

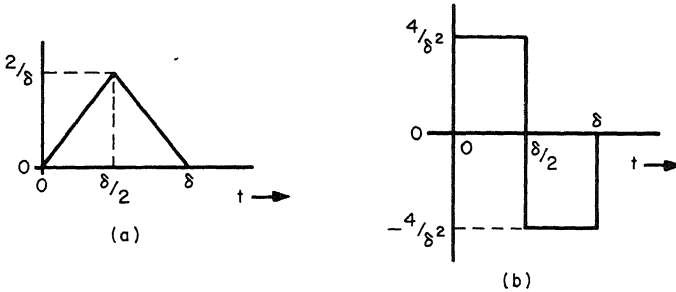


Fig. 2-5-8.

derivative is shown in Fig. 2-5-8b. The curve of Fig. 2-5-8b consists of two spikes of equal area, one spike being positive and the other negative. For this reason the derivative of the unit impulse is sometimes referred to as a unit doublet or a couple. The area under each spike is

$$\frac{4}{\delta^2} \times \frac{\delta}{2} = \frac{2}{\delta} \quad (2-5-11)$$

As δ approaches zero as a limit, the area under each spike increases without bound.

The wave shape of Fig. 2-5-8b can be visualized as being made up of three components and its transform is

$$\frac{4}{\delta^2 s} - \frac{8}{\delta^2 s} e^{-(\delta/2)s} + \frac{4}{\delta^2 s} e^{-\delta s} \quad (2-5-12)$$

As δ approaches zero, Eq. 2-5-12 approaches the transform of the unit doublet. If the limit, is taken as the equation stands, the result is the indeterminate form $0/0$, which can be evaluated by differentiating the numerator and denominator with respect to δ the proper number of times and then taking the limit:

$$\begin{aligned} \mathcal{L}[u_{+1}(t)] &= \lim_{\delta \rightarrow 0} \frac{4 - 8e^{-(\delta/2)s} + 4e^{-\delta s}}{\delta^2 s} = \frac{0}{0} \\ &= \lim_{\delta \rightarrow 0} \frac{4s e^{-(\delta/2)s} - 4s e^{-\delta s}}{2\delta s} = \frac{0}{0} \\ &= \lim_{\delta \rightarrow 0} \frac{-s e^{-(\delta/2)s} + 2s e^{-\delta s}}{1} = s \end{aligned} \quad (2-5-13)$$

The symbol $u_{+1}(t)$ stands for the unit doublet and s is its transform.

Since an infinite number of wave shapes can be used which in the limit lead to the unit impulse, it seems reasonable to suppose that there are also an infinite number of wave shapes which in the limit lead to the unit doublet. A few of these are shown in Fig. 2-5-9.

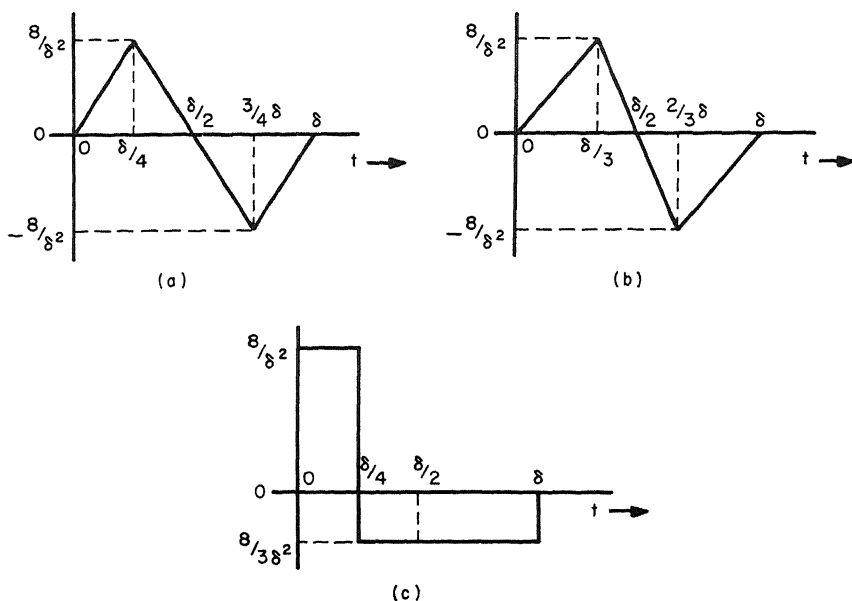


Fig. 2-5-9.

Example 2-5-7. The previous two examples introduce two members of a family of functions. Each successive member of the family can be obtained by differentiating the preceding member. Or in the reverse direction, each member of the family can be obtained by integrating the preceding member. The central member of this family is the unit impulse denoted by $u_0(t)$ which has 1 as its transform. If $u_0(t)$ is differentiated, the result is the unit doublet denoted by $u_{+1}(t)$, which has s as its transform. If $u_{+1}(t)$ is differentiated, the result is the unit triplet denoted by $u_{+2}(t)$ which has the transform s^2 . This line of reasoning can be extended to the general term as given by

$$\mathcal{L}[u_{+n}(t)] = s^n \quad (2-5-14)$$

In the reverse direction, the unit impulse $u_0(t)$ can be integrated and the result is the unit-step function denoted by $u_{-1}(t)$, which has the transform $1/s$. If $u_{-1}(t)$ is integrated, the result is the unit-ramp function denoted $u_{-2}(t)$, which has the transform $1/s^2$. This line of reasoning can be extended to the general term as given by

$$\mathcal{L}[u_{-n}(t)] = \frac{1}{s^n} \quad (2-5-15)$$

With the addition of the impulse family, all functions can be differentiated or integrated as many times as desired. When a function has a discontinuity and is differentiated, an impulse results. If the resulting function is differentiated, a doublet results, etc.

In certain situations, the use of the impulse family has another desirable consequence. If a function is defined from $-\infty$ to $+\infty$ and has a d-c value (more properly a constant value), and if the function is differentiated and the result in turn is integrated, the final result can differ from the original function by a constant unless an auxiliary step is included in reestablishing the d-c level. However, if the function is defined only for positive values of time and is equal to zero for negative time, integration is as exact as is differentiation if the impulse family is used. For example, the d-c level when considered for only positive values is actually a step function. When this step function is differentiated, an impulse is generated at $t = 0$. If the resulting function is integrated in turn, the integration of the impulse reestablishes the exact d-c level.

Finally, if a function is defined only for positive time and is equal to zero for negative time, the Laplace transform of a derivative of a function is equal to s times the transform of the function if the differentiation is performed in the generalized sense.

The transform pair

$$\mathcal{L}[f'(t)] = sF(s) - f(0) \quad (2-5-16)$$

applies to the situation in which differentiation is performed in the conventional sense. However, when differentiation is performed in the generalized sense, the transform pair becomes

$$\mathcal{L}[f'(t)] = sF(s) \quad (2-5-17)$$

In Eq. 2-5-17, $f'(t)$ contains an impulse if $f(0)$ has a value, and the equation is then identical with Eq. 2-5-16. This same reasoning process can be extended to derivatives of any order.

2-6. Additional Examples. The examples in this article are included to show the fascinating uses of some of these ideas.

Example 2-6-1. The transform for a train of unit impulses spaced T seconds apart, the first of which occurs at $t = 0$ (Fig. 2-6-1), can be obtained from

$$F_1(s) = 1 \quad (2-6-1)$$

whence $F(s)$ is

$$F(s) = F_1(s) \times \frac{1}{1 - e^{-Ts}} = \frac{1}{1 - e^{-Ts}} \quad (2-6-2)$$

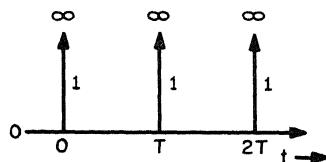
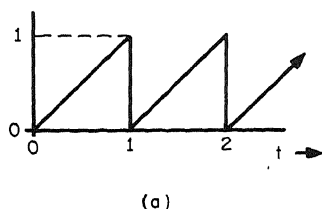


Fig. 2-6-1.

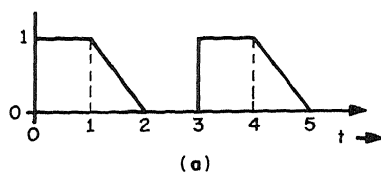
Likewise a train of unit doublets spaced T seconds apart, with the first occurring at $t = 0$, has the transform

$$F(s) = \frac{s}{1 - e^{-Ts}} \quad (2-6-3)$$

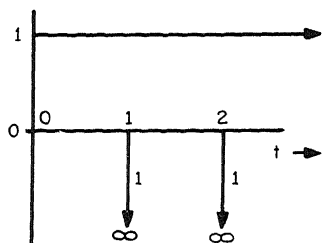
When a repeating wave shape is made up of segments having the nature of t^n , and if this repeating wave is differentiated $n + 1$ times, the result is a group of trains of impulses of various orders. When these impulse trains are integrated the proper number of times, the transform $F(s)$ of the original wave shape is obtained.



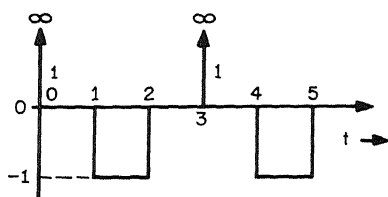
(a)



(a)



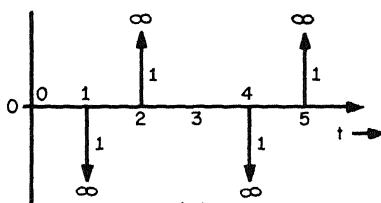
(b)



(b)



(c)



(c)

Fig. 2-6-2.

Fig. 2-6-3.

If the wave shape of Fig. 2-6-2a is differentiated once, the result is as shown in Fig. 2-6-2b. The train of impulses does not have an impulse at the origin. Therefore it can be regarded as the negative of the train shown in Fig. 2-6-1 delayed by 1 second. The transform for the train of impulses is

$$-\frac{\varepsilon^{-s}}{1 - \varepsilon^{-s}} \quad (2-6-4)$$

To continue until only impulses are left, the function of Fig. 2-6-2b must be differentiated again. When the train of impulses is differentiated, a train of doublets is obtained. This train of doublets must be integrated twice to recover its contribution to the $F(s)$ of the curve of Fig. 2-6-2a. Therefore, it is more convenient simply to hold back the impulse train and to differentiate the remainder of the curve, which yields an impulse at the origin as shown in Fig. 2-6-2c. Accordingly, the desired $F(s)$ is made up of the impulse of Fig. 2-6-2c integrated twice and the impulse train of Fig. 2-6-2b integrated once. Finally, $F(s)$ is obtained as

$$F(s) = \frac{1}{s^2} - \frac{\varepsilon^{-s}}{s(1 - \varepsilon^{-s})} \quad (2-6-5)$$

Example 2-6-2. The wave shape of Fig. 2-6-3a is differentiated once as shown in Fig. 2-6-3b. The resulting train of impulses is held back and the remaining curve is differentiated again as shown in Fig. 2-6-3c. The impulse train to be integrated once has the transform

$$\frac{1}{1 - e^{-3s}} \quad (2-6-6)$$

The impulse trains to be integrated twice have the transform

$$-\frac{e^{-s}}{1 - e^{-3s}} + \frac{e^{-2s}}{1 - e^{-3s}} \quad (2-6-7)$$

The transform of the original function is

$$\frac{1}{s(1 - e^{-3s})} + \frac{-e^{-s} + e^{-2s}}{s^2(1 - e^{-3s})} \quad (2-6-8)$$

Certain functions never disappear no matter how many times they are differentiated. For example, when the sine function is differentiated the successive results are +cosine, -sine, -cosine, and back to +sine. The method as explained in the previous two examples can be extended to functions of this type by taking advantage of the fact that the wave shape repeats after the function is differentiated a sufficient number of times.

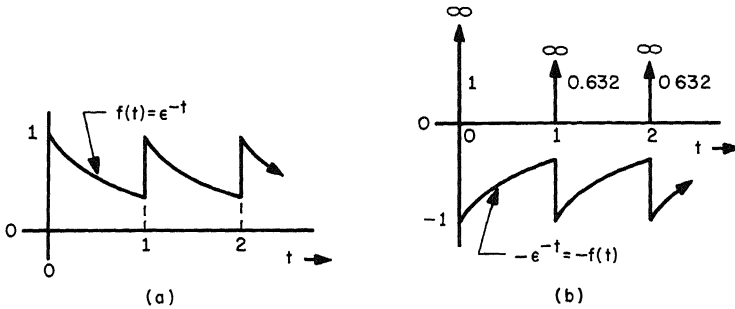


Fig. 2-6-4.

Example 2-6-3. This concept is best demonstrated by an example using the function of Fig. 2-6-4a. The derivative of the function is shown in Fig. 2-6-4b. When the functions of Fig. 2-6-4a and b are compared, the following equation can be written

$$\frac{d}{dt} [f(t)] = -f(t) + [\text{the impulses shown in Fig. 2-6-4b}] \quad (2-6-9)$$

This equation can be rewritten as

$$\frac{d}{dt} [f(t)] + f(t) = [\text{the impulses shown in Fig. 2-6-4b}] \quad (2-6-10)$$

The transform of this equation is

$$sF(s) + F(s) = 1 + \frac{0.632e^{-s}}{1 - e^{-3s}} \quad (2-6-11)$$

This can be solved for $F(s)$, giving

$$F(s) = \frac{1 - 0.368e^{-s}}{(s + 1)(1 - e^{-s})} \quad (2-6-12)$$

Techniques almost identical to these can also be used to obtain the Fourier series of repeating wave shapes.

2-7. Discussion of Function $F(s)$. The direct Laplace transform changes an $f(t)$ into an $F(s)$ which is a function of a complex variable $s = \sigma + j\omega$. The field of study involving functions of a complex variable is well developed, and many of the concepts used in this article and throughout the book are applications of some of the simpler concepts of this theory.

Most of the $F(s)$ functions of interest in control system theory are single-valued functions of s . That is, for a specific value of the complex variable s , $F(s)$ is uniquely determined and also is complex, having a real part and an imaginary part. In general,

$$F(s) = U(\sigma, \omega) + jV(\sigma, \omega)$$

where $U(\sigma, \omega)$ is the real part of $F(s)$ and $jV(\sigma, \omega)$ is the imaginary part.

A function $F(s)$ is said to be analytic at a point s if $F(s)$ has a unique derivative at each point in a region including the point s . If the definition for the derivative of a real function of a single variable is extended to functions of a complex variable, the derivative is given by

$$\frac{d[F(s)]}{ds} = \lim_{\Delta s \rightarrow 0} \frac{\Delta F(s)}{\Delta s} = \lim_{\Delta s \rightarrow 0} \frac{F(s + \Delta s) - F(s)}{\Delta s} \quad (2-7-1)$$

where $\Delta F(s) = \Delta U + j\Delta V$ and the limit is independent of the choice of Δs . The situation expressed by this equation is more complicated than the corresponding case for real functions in the sense that $\Delta s = \Delta\sigma + j\Delta\omega$ can approach zero along an infinite number of different paths. For instance, suppose $\Delta\omega = 0$ and $\Delta s = \Delta\sigma$. As Δs approaches zero in this manner, Eq. 2-7-1 becomes

$$\frac{d[F(s)]}{ds} = \lim_{\Delta\sigma \rightarrow 0} \left[\frac{\Delta U}{\Delta\sigma} + \frac{j\Delta V}{\Delta\sigma} \right] = \frac{\partial U}{\partial\sigma} + j\frac{\partial V}{\partial\sigma} \quad (2-7-2)$$

A second path along which Δs can approach zero is that for which $\Delta\sigma = 0$ and $\Delta s = j\Delta\omega$. As Δs approaches zero in this manner, Eq. 2-7-1 becomes

$$\frac{d[F(s)]}{ds} = \lim_{j\Delta\omega \rightarrow 0} \left[\frac{\Delta U}{j\Delta\omega} + \frac{j\Delta V}{j\Delta\omega} \right] = -j\frac{\partial U}{\partial\omega} + \frac{\partial V}{\partial\omega} \quad (2-7-3)$$

For the derivative to be unique, Eqs. 2-7-2 and 2-7-3 must be equal, as shown by

$$\frac{\partial U}{\partial\sigma} + j\frac{\partial V}{\partial\sigma} = \frac{\partial V}{\partial\omega} - j\frac{\partial U}{\partial\omega} \quad (2-7-4)$$

For two complex expressions to be equal, the real parts must be equal and so must the imaginary parts. This leads to the relationships

$$\frac{\partial U}{\partial \sigma} = \frac{\partial V}{\partial \omega} \text{ and } \frac{\partial V}{\partial \sigma} = -\frac{\partial U}{\partial \omega} \quad (2-7-5)$$

These equations are the well-known Cauchy-Riemann conditions. It can be shown that if Δs approaches zero along the two paths just discussed and the respective derivatives are equal, then Δs can approach zero along *any* path and the derivative is unique.

More generally, if both U and V and all four partial derivatives are continuous functions of s and satisfy the Cauchy-Riemann conditions at each point in a region including a specified point s , the function is analytic at the point s .

An example often given is

$$F(s) = s^2 = \sigma^2 - \omega^2 + j2\sigma\omega \quad (2-7-6)$$

In this case

$$\frac{\partial U}{\partial \sigma} = 2\sigma = \frac{\partial V}{\partial \omega} \text{ and } \frac{\partial V}{\partial \sigma} = 2\omega = -\frac{\partial U}{\partial \omega} \quad (2-7-7)$$

Therefore s^2 is analytic at every finite point in the s -plane and its derivative can be found to be $2s$ by use of either Eq. 2-7-2 or Eq. 2-7-3. In general, $F(s) = s^n$ also can be shown to be analytic at every finite point in the s -plane and its derivative is ns^{n-1} .

Examples of functions of s that are not analytic are $|s|$, the magnitude of s , and $\bar{s}$, the conjugate of s .

The point (or points) at which a function $F(s)$ is not analytic is called a singular point. The type of $F(s)$ function encountered most often in control systems is a ratio of two polynomials in s . A typical $F(s)$ given in factored form is

$$F(s) = \frac{K(s + z_1)(s + z_2)^2}{(s + p_1)(s + p_2)^2(s + p_3)^4} \quad (2-7-8)$$

This $F(s)$ has singular points at $s = -p_1$, $s = -p_2$, and $s = -p_3$. Actually, these values of s represent very specific types of singularities that are termed poles. A pole is defined in the following manner. If an $F(s)$ increases without bound as s approaches a value $-s_0$ and in such a manner that the function

$$G(s) = (s + s_0)^n F(s) \quad (2-7-9)$$

has a finite non-zero value at $-s_0$ for some positive integer n , then $-s_0$ is termed a pole of order n of $F(s)$. When n is equal to unity, the pole is called a simple pole. For the $F(s)$ of Eq. 2-7-8, $s = -p_1$ is a simple pole; $s = -p_2$ is a second-order pole; and $s = -p_3$ is a fourth-order pole.

A value of s that makes $F(s)$ equal to zero is termed a zero of the function. For the $F(s)$ of Eq. 2-7-8, $s = -z_1$ is a simple zero and $s = -z_2$ is a second-order zero.

If the point at infinity also is included, a function has as many zeros as it has poles, and vice versa. The example being considered has three finite zeros and seven finite poles. Hence there must be four zeros at infinity. The truth of this can be demonstrated by the fact that as $|s|$ approaches infinity, each of the z 's and p 's can eventually be ignored when compared with s . Therefore,

$$\lim_{|s| \rightarrow \infty} F(s) = \lim_{|s| \rightarrow \infty} \frac{Ks^3}{s^7} = \lim_{|s| \rightarrow \infty} \frac{K}{s^4} \quad (2-7-10)$$

and the fourth-order zero at infinity is displayed.

As a second example, the function

$$F(s) = \frac{K(s + z_1)(s + z_2)}{(s + p_1)(s + p_2)(s + p_3)} \quad (2-7-11)$$

is considered. It is helpful to locate the poles and zeros on a sketch of the complex s -plane as shown in Fig. 2-7-1. The locations of the zeros are

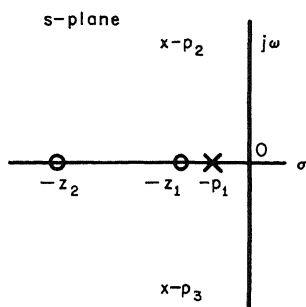


Fig. 2-7-1.

shown by circles and those of the poles by x symbols. Except for the K factor of Eq. 2-7-11, the pole-zero configuration completely specifies the $F(s)$ function.

If the magnitude of the $F(s)$ function is considered as equal to the distance perpendicular to the s -plane of Fig. 2-7-1, then as s takes on all possible values, the plot of $|F(s)|$ forms a three-dimensional surface. For instance, as s approaches one of the zeros, say $-z_1$, $F(s)$ approaches the value zero. Likewise if s approaches one of the poles, say $-p_1$, $|F(s)|$ increases in magnitude without bound. The geometry of $|F(s)|$ in the vicinity of a pole suggests a pole.

A rather crude analogy to the shape of an $|F(s)|$ surface is that of a thin rubber membrane stretched taut over a large tabletop. If $|F(s)|$ has a zero at $s = \infty$, the outer portions of the membrane are clamped to the tabletop. At the locations of the finite zeros, the rubber membrane is held to the

tabletop with tacks of zero diameter. As a practical consideration, each of the tacks should have the same small diameter. At the poles, the membrane should be pushed upward to an infinite height by rods of zero diameter. As a practical consideration, the rods should all have the same small diameter and be pushed upward to the same height. The analogy is crude, but the shape of the rubber membrane represents in a general way the $|F(s)|$ surface.

Let it be supposed that some $f(t)$ is given and the direct Laplace integral is used to find the corresponding $F(s)$. As a specific example, the $F(s)$ of Eq. 2-7-11 which has the pole-zero configuration shown in Fig. 2-7-1 is considered. As has been stressed, the direct Laplace integral is an improper integral and converges only if σ (the real part of s) is larger than σ_0 . In terms of the poles and zeros of an $F(s)$, σ_0 is the real part of the pole located farthest to the right in the complex s -plane. For the configuration shown in Fig. 2-7-1, $\sigma_0 = -p_1$.

For present purposes, the direct Laplace integral is written with s replaced by its real part σ .

$$F(\sigma) = \int_0^{\infty} f(t)e^{-\sigma t} dt \quad (2-7-12)$$

The integral converges along the portion of the real axis to the right of σ_0 .

In the theory of functions of a complex variable, there is a concept known as analytic continuation which is helpful for the present discussion. If a function is known to be analytic in some region and its values are known along some arc in this region, the function is uniquely determined throughout the region. This concept can be applied to the present discussion in the following manner. $F(s)$ can be found by using the direct Laplace transform as discussed earlier in this chapter. $F(\sigma)$ can be found by use of Eq. 2-7-12 and thus $F(\sigma)$ is certainly valid along an arc which is the real axis to the right of σ_0 . $F(s)$ and $F(\sigma)$ are identical along the σ axis to the right of σ_0 . Therefore $F(s)$ can be considered as the analytic continuation of $F(\sigma)$ throughout the entire s -plane except at the points where $F(s)$ is not analytic.

2-8. The Inverse Transformation. The inverse Laplace transformation can be performed by using the complex inversion integral given by

$$\mathcal{L}^{-1}[F(s)] = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} F(s)e^{st} ds = f(t) \quad (2-8-1)$$

The path of integration indicated by this equation is a path parallel to and c units away from the $j\omega$ axis as shown in Fig. 2-8-1. The quantity c is chosen so that the path of integration is to the right of all the poles and branch points in the complex s -plane.

Use of the complex inversion integral demands more knowledge of the theory of functions of a complex variable than seems appropriate in a book

of this type. Therefore, the method to be employed is to use the table of transform pairs in the direction opposite to that in which the table was developed. Sometimes the $F(s)$ function must be manipulated into a form that is in the table. Arts. 2-9 and 2-10 are concerned with this problem.

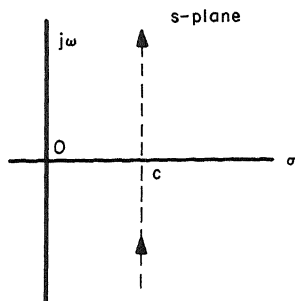


Fig. 2-8-1.

2-9. Partial Fraction Expansion. Examples are used to explain the concepts involved.

Example 2-9-1. The $F(s)$ for the first example is

$$F(s) = \frac{(s+2)(s+5)}{(s+1)(s+3)(s+4)} = \frac{K_{-1}}{s+1} + \frac{K_{-3}}{s+3} + \frac{K_{-4}}{s+4} \quad (2-9-1)$$

The right side of this equation is the partial-fraction expansion. The coefficients of simple poles such as those of this example are given the name residues.

When Eq. 2-9-1 is multiplied by $s+1$, the result is

$$\frac{(s+2)(s+5)}{(s+3)(s+4)} = K_{-1} + \frac{K_{-3}(s+1)}{s+3} + \frac{K_{-4}(s+1)}{s+4} \quad (2-9-2)$$

When s is assigned the value $s = -1$, the residue K_{-1} is found to be

$$K_{-1} = \frac{(-1+2)(-1+5)}{(-1+3)(-1+4)} = \frac{2}{3} \quad (2-9-3)$$

When Eq. 2-9-1 is multiplied by $s+3$, the result is

$$\frac{(s+2)(s+5)}{(s+1)(s+4)} = \frac{K_{-1}(s+3)}{s+1} + K_{-3} + \frac{K_{-4}(s+3)}{s+4} \quad (2-9-4)$$

When s is assigned the value -3 , the residue K_{-3} is found to be

$$K_{-3} = \frac{(-3+2)(-3+5)}{(-3+1)(-3+4)} = 1 \quad (2-9-5)$$

In a similar manner, K_{-4} can be found as

$$K_{-4} = \frac{(-4+2)(-4+5)}{(-4+1)(-4+3)} = -\frac{2}{3} \quad (2-9-6)$$

When these residues are substituted in Eq. 2-9-1, the resulting partial fraction expansion is

$$\frac{(s+2)(s+5)}{(s+1)(s+3)(s+4)} = \frac{2/3}{s+1} + \frac{1}{s+3} - \frac{2/3}{s+4} \quad (2-9-7)$$

The inverse transform can now be taken as

$$\mathcal{L}^{-1} \left[\frac{(s+2)(s+5)}{(s+1)(s+3)(s+4)} \right] = \frac{2}{3} e^{-t} + e^{-3t} - \frac{2}{3} e^{-4t} \quad (2-9-8)$$

A rule of thumb that is convenient to remember can be stated thus: When a residue such as K_{-1} in the pole at $s = -1$ is to be evaluated, the factor $s+1$ is "covered up" and s throughout the remaining portion of the expression is assigned the value that makes the "covered-up" term equal to zero. This technique is a simplified version of multiplying both sides of an equation by the factor. By use of this procedure, the partial fraction expansion can be written by inspection.

Example 2-9-2. As a second example, the following $F(s)$ is considered:

$$F(s) = \frac{s^3 + 5s^2 + 6s - 1}{(s+1)(s+2)} \quad (2-9-9)$$

This $F(s)$ is an improper fraction in that the numerator is of higher degree in s than the denominator. One procedure that can be followed is division of the denominator into the numerator until the remainder is a proper fraction.

$$\begin{array}{r} s+2 \\ s^2+3s+2 \overline{) s^3+5s^2+6s-1} \\ \underline{s^3+3s^2+2s} \\ 2s^2+4s-1 \\ \underline{2s^2+6s+4} \\ -2s-5 \end{array} \quad (2-9-10)$$

The $F(s)$ function can be written as

$$\frac{s^3 + 5s^2 + 6s - 1}{(s+1)(s+2)} = s + 2 + \frac{-2s - 5}{(s+1)(s+2)} \quad (2-9-11)$$

or as

$$\frac{s^3 + 5s^2 + 6s - 1}{(s+1)(s+2)} = s + 2 + \frac{K_{-1}}{(s+1)} + \frac{K_{-2}}{(s+2)} \quad (2-9-12)$$

The residue K_{-1} can be found as before by multiplying both sides of the equation by $s+1$:

$$\frac{s^3 + 5s^2 + 6s - 1}{s+2} = (s+2)(s+1) + K_{-1} + \frac{K_{-2}(s+1)}{s+2} \quad (2-9-13)$$

and letting $s = -1$ in the result; thus,

$$K_{-1} = \frac{(-1)^3 + 5(-1)^2 + 6(-1) - 1}{-1 + 2} = -3 \quad (2-9-14)$$

In a similar manner, K_{-2} is found as

$$K_{-2} = \left. \frac{s^3 + 5s^2 + 6s - 1}{s+1} \right|_{s=-2} = 1 \quad (2-9-15)$$

Upon substitution, Eq. 2-9-12 becomes

$$\frac{s^3 + 5s^2 + 6s - 1}{(s+1)(s+2)} = s + 2 - \frac{3}{s+1} + \frac{1}{s+2} \quad (2-9-16)$$

and the inverse transformation is

$$\mathcal{L}^{-1}\left[\frac{s^3 + 5s^2 + 6s - 1}{(s+1)(s+2)}\right] = u_{-1}(t) + 2u_0(t) - 3e^{-t} + e^{-2t} \quad (2-9-17)$$

As in the previous example, the procedure for determining K_{-1} and K_{-2} can be condensed by using the "cover-up" technique.

In most cases, it is probably easier to expand only the proper fraction portion of an equation such as Eq. 2-9-11 as indicated by

$$\frac{-2s - 5}{(s+1)(s+2)} = \frac{K_{-1}}{s+1} + \frac{K_{-2}}{s+2} \quad (2-9-18)$$

In this equation, the K 's are found as

$$K_{-1} = \left. \frac{-2s - 5}{s+2} \right|_{s=-1} = -3 \quad (2-9-19)$$

$$K_{-2} = \left. \frac{-2s - 5}{s+1} \right|_{s=-2} = 1 \quad (2-9-20)$$

The work implied by Eqs. 2-9-19 and 2-9-20 is somewhat less than that of Eqs. 2-9-14 and 2-9-15 respectively.

Example 2-9-3. As a third example, an $F(s)$ with a second-order pole is examined:

$$F(s) = \frac{(s+2)(s+5)}{(s+1)(s+3)^2} = \frac{K_{-1}}{s+1} + \frac{K_{-3}}{s+3} + \frac{A}{(s+3)^2} \quad (2-9-21)$$

The residue K_{-1} can be found by the "cover-up" technique as

$$K_{-1} = \left. \frac{(s+2)(s+5)}{(s+3)^2} \right|_{s=-1} = 1 \quad (2-9-22)$$

The term A can be found by multiplying Eq. 2-9-21 by $(s+3)^2$:

$$\frac{(s+2)(s+5)}{s+1} = \frac{K_{-1}(s+3)^2}{s+1} + K_{-3}(s+3) + A \quad (2-9-23)$$

and by letting $s = -3$ in the result.

$$A = \left. \frac{(s+2)(s+5)}{s+1} \right|_{s=-3} = 1 \quad (2-9-24)$$

In order to obtain the residue K_{-3} , Eq. 2-9-23 can be differentiated with respect to s and, in the result, s given the value $s = -3$; thus,

$$K_{-3} = \left. \frac{d}{ds} \frac{(s+2)(s+5)}{s+1} \right|_{s=-3} = 0 \quad (2-9-25)$$

Upon substitution, Eq. 2-9-21 becomes

$$\frac{(s+2)(s+5)}{(s+1)(s+3)^2} = \frac{1}{s+1} + \frac{1}{(s+3)^2} \quad (2-9-26)$$

and the inverse transform is

$$\mathcal{L}^{-1} \left[\frac{(s+2)(s+5)}{(s+1)(s+3)^2} \right] = e^{-t} + te^{-3t} \quad (2-9-27)$$

2-10. Heaviside Expansion Theorem. The Heaviside expansion theorem is another method for obtaining the partial fraction expansion. The theorem is presented here for the case of simple poles. The $F(s)$ function is a ratio of two polynomials as

$$F(s) = \frac{N(s)}{D(s)} \quad (2-10-1)$$

It is assumed that $s + p_1$ is one of the factors of $D(s)$ and that the product of all other factors in $D(s)$ is given by $G(s)$. In other words, $D(s)$ is

$$D(s) = (s + p_1)G(s) \quad (2-10-2)$$

$F(s)$ can now be written as

$$F(s) = \frac{N(s)}{(s + p_1)G(s)} \quad (2-10-3)$$

If this $F(s)$ is expanded in partial fractions, the residue K_{-p_1} in the pole at $s = -p_1$, as found by using the "cover-up" technique, is

$$K_{-p_1} = \frac{N(-p_1)}{G(-p_1)} \quad (2-10-4)$$

The Heaviside expansion theorem achieves this same result by an alternate procedure. The residue K_{-p_1} is found from

$$K_{-p_1} = \frac{N(s)}{\frac{d}{ds}[D(s)]} \Big|_{s=-p_1} \quad (2-10-5)$$

K_{-p_1} of Eqs. 2-10-4 and 2-10-5 is the same, as can be seen from the following:

$$\begin{aligned} K_{-p_1} &= \frac{N(s)}{\frac{d}{ds}[D(s)]} \Big|_{s=-p_1} = \frac{N(s)}{\frac{d}{ds}[(s + p_1)G(s)]} \Big|_{s=-p_1} \\ &= \frac{N(s)}{(s + p_1)G'(s) + G(s)} \Big|_{s=-p_1} = \frac{N(-p_1)}{G(-p_1)} \end{aligned} \quad (2-10-6)$$

Example 2-10-1. The wave shape shown in Fig. 2-6-4a is considered again. The $F(s)$ corresponding to this wave shape is found in Art. 2-6 to be

$$F(s) = \frac{1 - 0.368e^{-s}}{(s+1)(1 - e^{-s})} \quad (2-10-7)$$

At first glance, it may seem that $s = -1$ is a pole of this $F(s)$. However, for $s = -1$, $F(s)$ has the indeterminate form $0/0$ which when evaluated is found to be finite. Therefore $s = -1$ is not a pole of $F(s)$.

The poles of $F(s)$ are the zeros of the factor $1 - e^{-s}$, and thus are the roots of the equation

$$1 - e^{-s} = 0 \quad (2-10-8)$$

which can be recast as

$$e^{-s} = 1 \quad (2-10-9)$$

When s is replaced by $s = \sigma + j\omega$, this equation becomes

$$e^{-(\sigma + j\omega)} = e^{-\sigma}[\cos \omega - j \sin \omega] = 1 \quad (2-10-10)$$

From this equation, it follows that

$$\sigma = 0; \omega = \pm 2n\pi \quad (n = 0, 1, 2, \dots) \quad (2-10-11)$$

Therefore the poles of $F(s)$ are

$$s = \pm j2n\pi \quad (n = 0, 1, 2, \dots) \quad (2-10-12)$$

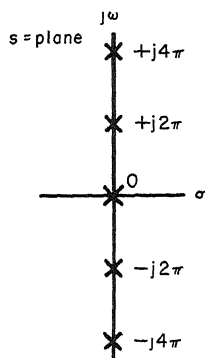


Fig. 2-10-1.

Fig. 2-10-1 shows a portion of the s -plane, and the location of a few of the poles is shown.

The Heaviside expansion theorem can be used to find the residues. The use of the theorem in this case is an extension of the theorem from the form in which it is shown to be valid in Eq. 2-10-6:

$$\begin{aligned} K_{+j2n\pi} &= \frac{1 - 0.368e^{-s}}{\frac{d}{ds}[(s+1)(1-e^{-s})]} \bigg|_{s=j2n\pi} \\ &= \frac{1 - 0.368e^{-s}}{(s+1)e^{-s} + (1-e^{-s})} \bigg|_{s=j2n\pi} \\ &= \frac{1 - 0.368}{j2n\pi + 1} = \frac{0.632}{1 + j2n\pi} \end{aligned} \quad (2-10-13)$$

A few of the terms of $F(s)$ can now be written:

$$\begin{aligned} F(s) &= \dots + \frac{0.004 - j0.0503}{s - j4\pi} + \frac{0.0156 - j0.098}{s - j2\pi} + \frac{0.632}{s} \\ &\quad + \frac{0.0156 + j0.098}{s + j2\pi} + \frac{0.004 + j0.0503}{s + j4\pi} + \dots \end{aligned} \quad (2-10-14)$$

The terms of $F(s)$ can be combined in pairs as

$$F(s) = \frac{0.632}{s} + \frac{0.0312s + 1.23}{s^2 + (2\pi)^2} + \frac{0.008s + 1.263}{s^2 + (4\pi)^2} + \dots \quad (2-10-15)$$

The inverse transformation can now be performed and the result is

$$\begin{aligned} \mathcal{L}^{-1} \left[\frac{1 - 0.368e^{-s}}{(s+1)(1-e^{-s})} \right] &= 0.632 + 0.196 \sin 2\pi t + 0.0312 \cos 2\pi t \\ &\quad + 0.1006 \sin 4\pi t + 0.008 \cos 4\pi t + \dots \end{aligned} \quad (2-10-16)$$

Eq. 2-10-16 shows a few terms of the infinite series that is the well-known Fourier-series development of the wave shape shown in Fig. 2-6-4a. If this wave shape represents some forcing quantity that is applied to a linear physical system and the response of the system is obtained by Laplace transform methods, the response function will contain the transient response as well as the steady-state response normally represented by the Fourier series.

2-11. Final-Value Theorem. A brief development of the final-value theorem shows how the theorem can be used. Of equal importance, the development shows the situation in which the theorem cannot be used. The theorem is developed by using a specific $F(s)$ of the form

$$F(s) = \frac{K(s + z_1)(s + z_2)}{s(s + p_1)(s + p_2)} = \frac{K_0}{s} + \frac{K_{-p_1}}{s + p_1} + \frac{K_{-p_2}}{s + p_2} \quad (2-11-1)$$

The inverse transform of this equation is

$$\mathcal{L}^{-1}[F(s)] = K_0 + K_{-p_1}e^{-p_1 t} + K_{-p_2}e^{-p_2 t} \quad (2-11-2)$$

The final value of $f(t)$ is K_0 , which can be obtained by letting $t \rightarrow \infty$ in Eq. 2-11-2. The coefficient K_0 is obtained from $F(s)$ by multiplying $F(s)$ by s and then letting s approach zero in the product. Therefore, the final-value theorem can be stated as

$$\lim_{t \rightarrow \infty} [f(t)] = \lim_{s \rightarrow 0} s[F(s)] \quad (2-11-3)$$

where $F(s)$ is the transform of $f(t)$. The only time $\lim_{s \rightarrow 0} s[F(s)]$ can have a finite non-zero value is when the denominator of $F(s)$ has s as a factor.

The final-value theorem is valid if $\lim_{t \rightarrow \infty} f(t)$ exists. An examination of Eq. 2-11-2 indicates that if p_1 or p_2 or both are negative, the final value of $f(t)$ is infinite, whereas the theorem gives a finite value. Therefore, the theorem can be applied to an $F(s)$ that has all its poles in the left-half s -plane, except possibly for a single pole at the origin.

2-12. Initial-Value Theorem. The initial-value theorem can best be presented by using the equation for the Laplace transform of a derivative of a function in which differentiation is performed in the conventional manner:

$$\int_0^{\infty} f'(t)e^{-st} dt = sF(s) - f(0) \quad (2-12-1)$$

The Laplace integral converges in a region in the s -plane to the right of $s = \sigma_0$. The limit as $s \rightarrow \infty$ can be taken on both sides of Eq. 2-12-1:

$$\lim_{s \rightarrow \infty} \int_0^{\infty} f'(t)e^{-st} dt = \lim_{s \rightarrow +\infty} sF(s) - f(0) \quad (2-12-2)$$

The limiting process yields

$$0 = \lim_{s \rightarrow \infty} sF(s) - f(0) \quad (2-12-3)$$

which can be written as

$$\lim_{t \rightarrow 0} f(t) = \lim_{s \rightarrow \infty} s[F(s)] \quad (2-12-4)$$

In essence, this equation is a statement of the initial-value theorem.

If the Laplace transform of the derivative of a function is written on the assumption that differentiation is performed in the generalized manner, the equation is

$$\int_0^{\infty} f'(t) e^{-st} dt = s[F(s)] \quad (2-12-5)$$

The limiting process

$$\lim_{s \rightarrow +\infty} \int_0^{\infty} f'(t) e^{-st} dt = \lim_{s \rightarrow \infty} s[F(s)] \quad (2-12-6)$$

yields

$$f(0) = \lim_{s \rightarrow \infty} s[F(s)] \quad (2-12-7)$$

Since $f(t)$ is zero for negative time, the derivative of an $f(t)$ that has an initial value $f(0)$ contains an impulse with area equal to $f(0)$. Therefore the $f(0)$ in Eq. 2-12-7 comes from the integral of Eq. 2-12-6.

Example 2-12-1. As an example of the use of the initial-value theorem, the $F(s)$ of Eq. 2-9-1 is used

$$F(s) = \frac{(s+2)(s+5)}{(s+1)(s+3)(s+4)} \quad (2-12-8)$$

The initial value of $f(t)$ is found as

$$f(0) = \lim_{s \rightarrow \infty} \frac{s(s+2)(s+5)}{(s+1)(s+3)(s+4)} = 1 \quad (2-12-9)$$

The inverse transform for this $F(s)$ is found as Eq. 2-9-8 which is repeated for convenience as

$$f(t) = \frac{2}{3} e^{-t} + e^{-3t} - \frac{2}{3} e^{-4t} \quad (2-12-10)$$

The use of the initial-value theorem provides a partial check on the accuracy of the solution.

An examination of the use of this theorem shows that the $f(0)$ is of finite non-zero value only if the degree of the numerator is one less than the degree of the denominator.

Example 2-12-2. As a second example, the function

$$F(s) = \frac{1}{(s+9)^3} \quad (2-12-11)$$

is used. The initial value is found to be

$$f(0) = \lim_{s \rightarrow \infty} \frac{s}{(s+9)^3} = 0 \quad (2-12-12)$$

The transform of the first derivative of $f(t)$ is

$$\mathcal{L}[f'(t)] = s \mathcal{L}[f(t)] - f(0) \quad (2-12-13)$$

where the differentiation is displayed in the conventional manner. In this case $f(0)$ has just been shown to be zero. Therefore the transform of the derivative is

$$\mathcal{L}[f'(t)] = \frac{s}{(s+9)^3} \quad (2-12-14)$$

and the initial value of the $f'(t)$ can be found as

$$f'(0) = \lim_{s \rightarrow \infty} \frac{s^2}{(s+9)^3} = 0 \quad (2-12-15)$$

The transform of the second derivative is

$$\mathcal{L}[f''(t)] = s^2 \mathcal{L}[f(t)] - sf(0) - f'(0) \quad (2-12-16)$$

Therefore, the transform of the second derivative is

$$\mathcal{L}[f''(t)] = \frac{s^2}{(s+9)^3} \quad (2-12-17)$$

and the initial value of the $f''(t)$ can be found as

$$f''(0) = \lim_{s \rightarrow \infty} \frac{s^3}{(s+9)^3} = 1 \quad (2-12-18)$$

In the manner just demonstrated, the nature of $f(t)$ and its derivatives at $t = 0$ can be investigated from a given $F(s)$.

PROBLEMS

2-1. Verify that

$$\mathcal{L}[f'''(t)] = s^3 \mathcal{L}[f(t)] - s^2 f(0) - sf'(0) - f''(0)$$

2-2. Verify that

$$\mathcal{L}[\cos \omega t] = \frac{s}{s^2 + \omega^2}$$

by some method different from that used in Example 2-4-5.

2-3. Verify that

$$(a) \mathcal{L}[\cosh bt] = \frac{s}{s^2 - b^2}$$

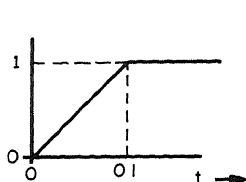
$$(b) \mathcal{L}[\sinh bt] = \frac{b}{s^2 - b^2}$$

$$(c) \mathcal{L}[t] = \frac{1}{s^2}$$

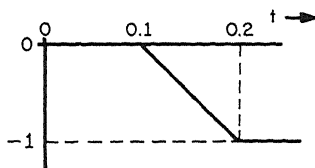
$$(d) \mathcal{L}[t^n] = \frac{n!}{s^{n+1}}$$

$$(e) \mathcal{L}[e^{-at}t^n] = \frac{n!}{(s+a)^{n+1}}$$

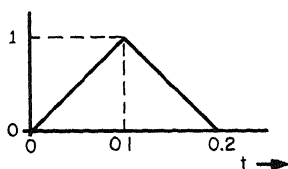
2-4. Check the items in the transform table in the Appendix as assigned.



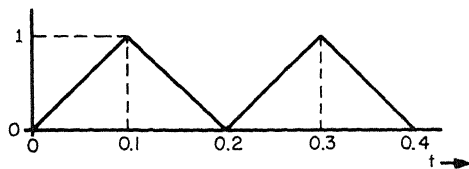
(a)



(b)



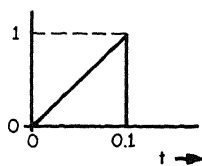
(c)



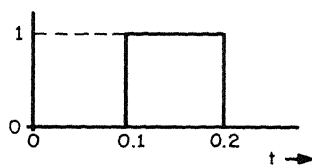
(d)

Fig. 2-P-5.

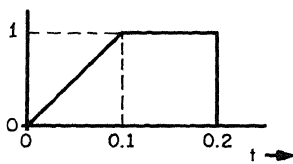
2-5. Determine the transform for the wave shapes in Fig. 2-P-5. Note that the wave form in Fig. 2-P-5c can be obtained by adding the wave forms of Fig. 2-P-5a and b.



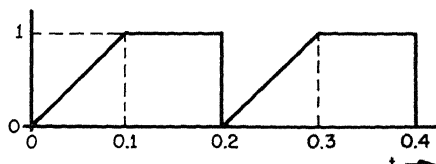
(a)



(b)



(c)



(d)

Fig. 2-P-6.

2-6. Determine the transform for the wave shapes in Fig. 2-P-6. Note that the wave form in Fig. 2-P-6c can be obtained by adding the wave forms of Fig. 2-P-6a and b.

2-7. Determine the transforms of the repeated wave shapes shown in Fig. 2-P-7 by using the methods developed in Art. 2-5.

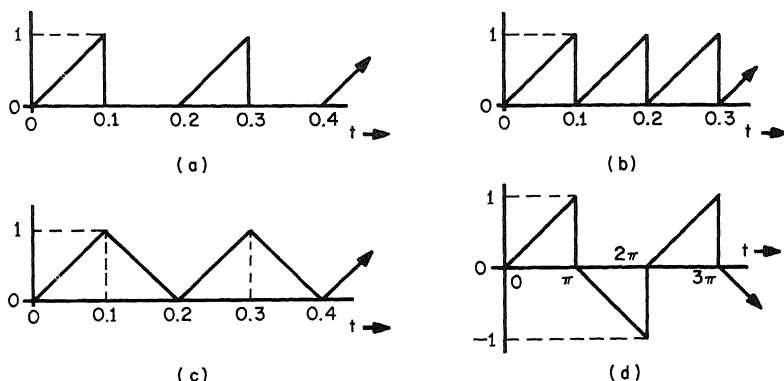


Fig. 2-P-7.

2-8. Start with the transform determined for the wave shape shown in Fig. 2-P-7c and perform the inverse transform, thus obtaining the Fourier series. Carry the solution out to three terms.

2-9. Start with the wave shape of Fig. 2-5-6a, develop the transform for the wave shape, and in the result let $\delta \rightarrow 0$, thus obtaining the transform for $u_0(t)$.

2-10. Repeat Prob. 2-8 except use the wave shape of Fig. 2-5-6b.

2-11. Start with the wave shape of Fig. 2-5-9a, develop the transform for the wave shape and in the result let $\delta \rightarrow 0$, thus obtaining the transform for $u_{+1}(t)$.

2-12. Repeat Prob. 2-11 except use the wave shape of Fig. 2-5-9b.

2-13. Repeat Prob. 2-11 except use the wave shape of Fig. 2-5-9c.

2-14. Repeat Prob. 2-P-7 except obtain the transforms of the various wave shapes by the methods developed in Art. 2-6.

2-15. Start with the transform determined for the wave shape shown in Fig. 2-P-7d and perform the inverse transform, thus obtaining the Fourier series. Carry the solution out to three terms.

2-16. Determine the transforms of the repeating wave shapes shown in Fig. 2-P-16 by using the methods outlined in Art. 2-6 and check by using the methods of Art. 2-5.

2-17. Repeat Prob. 2-15 except use the wave form shown in Fig. 2-P-16a.

2-18. Repeat Prob. 2-15 except use the wave form shown in Fig. 2-P-16b.

2-19. Repeat Prob. 2-15 except use the wave form shown in Fig. 2-P-16c.

2-20. Find the inverse transform $f(t)$ corresponding to each of the following $F(s)$ by using the "cover-up" technique:

$$(a) \frac{1}{s(s+2)(s+4)(s+6)}$$

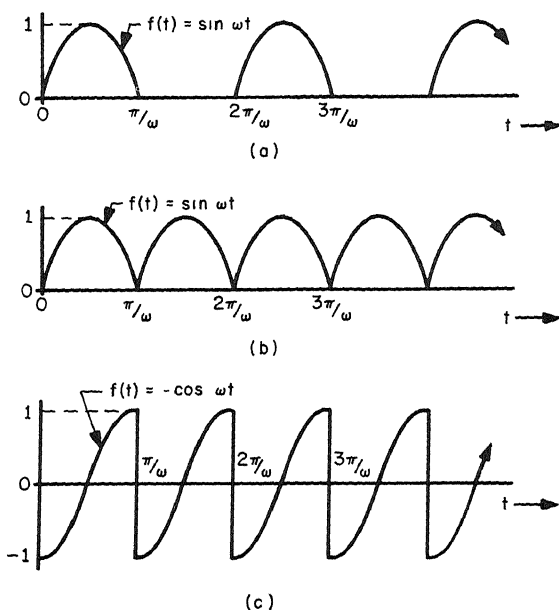


Fig. 2-P-16.

$$(b) \frac{s+1}{(s+2)(s+4)(s+7)}$$

$$(c) \frac{(s+1)(s+10)}{(s+2)(s+4)(s+7)}$$

$$(d) \frac{(s+1)(s+10)}{s(s+2)(s+4)(s+7)}$$

2-21. Repeat Prob. 2-20 except use Heaviside's expansion theorem to obtain each $f(t)$.

2-22. Find the inverse transform $f(t)$ corresponding to each of the following $F(s)$:

$$(a) \frac{(s+1)(s+6)(s+10)}{(s+5)(s+8)}$$

$$(b) \frac{(s+1)(s+6)(s+10)}{(s+5)(s+8)^2}$$

$$(c) \frac{s+3}{s(s+1)^2(s+4)^2}$$

$$(d) \frac{1}{s^4+1}$$

$$(e) \frac{1}{s^8+1}$$

2-23. Use the initial-value and final-value theorems to investigate the $f(t)$ functions that correspond with the $F(s)$ functions as given in Prob. 2-20. Check the information obtained by using these theorems with the $f(t)$ functions obtained by working Prob. 2-20.

2-24. Repeat Prob. 2-23 except use the $F(s)$ functions from Prob. 2-22.

2-25. Use the transform table in the Appendix and obtain the inverse transform of the following functions:

$$(a) \frac{16.6s + 4166}{s[s^2 + 250s + 10^5]}$$

$$(b) \frac{5,000}{s[s^2 + 250s + 10^5]}$$

$$(c) \frac{0.25 \times 10^6 s + 20 \times 10^7}{s[s^2 + 5.8 \times 10^4 s + 4 \times 10^7]}$$

$$(d) \frac{0.274s^2 + 250s + 38,900}{[s + 350][s^2 + (377)^2]}$$

$$(e) \frac{-571s^2}{[(s + 80)^2 + (890)^2][(s + 198.5)^2 + (84,500)^2]}$$

$$(f) \frac{-0.331 \times 10^4 s^3 + 1.34 \times 10^6 s^2 + 2.72 \times 10^8 s + 14.1 \times 10^{10}}{s[s^2 + (377)^2][s^2 + 200s + 6 \times 10^4]}$$

2-26. A train of impulses spaced T seconds apart is modulated by the function ε^{-at} . That is, the first impulse has a value of unity, the second impulse the value ε^{-aT} , etc. Obtain the transform of this impulse train as an infinite series and then obtain the sum function for the series. A change of variable $z = \varepsilon^{sT}$ yields the z -transform for the ε^{-at} function. Obtain this z -transform and label the answer $F(z)$.

2-27. Repeat Prob. 2-26 except modulate the train of impulses by the function t .

2-28. Repeat Prob. 2-26 except modulate the train of impulses by the function $\sin \omega t$.

3

ELECTRIC CIRCUITS

3-1. Introduction. Chapter 1 outlines in qualitative terms the basic concepts of control systems without attempting any quantitative analysis. Because control system theory encompasses electrical, mechanical, hydraulic, and pneumatic systems, among others, this chapter and the next two are concerned with the basic physical laws useful in the theory. From these laws, the equations describing the operation of the systems can be formulated and performance of the systems can be determined.

Chapters 3, 4, and 5 demonstrate that the analysis of all linear systems follows essentially the same pattern regardless of the nature of the system.

This book assumes a background in analytical mechanics and in electric circuit theory. If the reader lacks such a background, he should consult any of the standard books in these areas.

3-2. Electric Elements. The three basic elements used in electric circuits are the resistor, inductor, and capacitor possessing the qualities of resistance, R , inductance, L , and capacitance, C . It is assumed that these elements are constants independent of either the voltage or the current and the following relationships between current and voltage can be written:

For the resistor,

$$e_R = iR \quad \text{or} \quad i = \frac{e_R}{R} \quad (3-2-1)$$

For the inductor,

$$e_L = L \frac{di}{dt} \quad \text{or} \quad i = \frac{1}{L} \int_0^t e_L dt + i(0) \quad (3-2-2)$$

For the capacitor,

$$e_c = \frac{1}{C} \int_0^t i dt + e(0) \quad \text{or} \quad i = C \frac{de_c}{dt} \quad (3-2-3)$$

The symbols used to represent the R , L , and C are shown in Fig. 3-2-1. For each element, the $+$ and $-$ signs indicate the polarity reference for

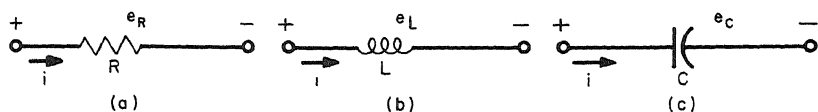


Fig. 3-2-1.

the voltage and the arrow indicates the reference direction for the current.

A fourth electric element, represented by the symbol M , is mutual inductance, which is subsequently discussed as needed at appropriate points in the book.

3-3. Kirchhoff's Voltage Law. Kirchhoff's voltage law states that the algebraic sum of the voltage drops around a closed path is zero. The law can also be expressed by stating that the sum of the voltage drops is equal to the sum of the voltage rises around a closed path.

In the general case, a number of closed paths must be transversed and a number of equations based on Kirchhoff's voltage law must be written. For the present purposes a simple example is chosen to demonstrate the manner in which Laplace transforms can be used to solve simple circuit problems.

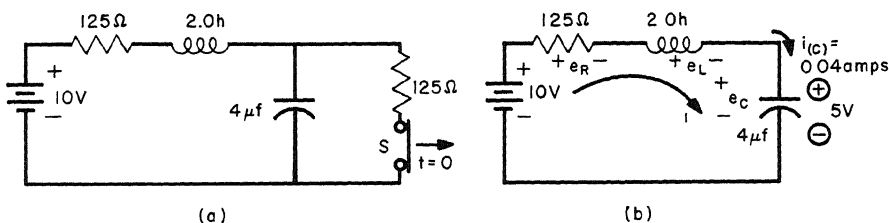


Fig. 3-3-1.

Example 3-3-1. The circuit shown in Fig. 3-3-1a is in steady state with switch S closed. S is then opened at $t = 0$. In steady state before $t = 0$, the inductor behaves as if it were a short circuit and the capacitor an open circuit. The current in the inductor is equal to 10 volts divided by 250 ohms (0.04 ampere). The voltage on the capacitor is equal to 5 volts as the two equal resistors divide the 10 volts equally between them.

At $t = 0$ the initial current in the inductor is $i(0) = 0.04$ ampere and the initial voltage on the capacitor is 5 volts. The circuit after $t = 0$ is shown in Fig. 3-3-1b. The $\oplus$ and $\ominus$ marks on the figure indicate the polarity of the initial voltage on the capacitor.

The arbitrary direction for the current i is chosen and Kirchhoff's voltage law is written as

$$10 = 125i + 2.0 \frac{di}{dt} + \frac{1}{4 \times 10^{-6}} \int_0^t i dt + 5 \quad (3-3-1)$$

The Laplace transform of this equation is

$$\frac{10}{s} = 125I(s) + 2[sI(s) - 0.04] + \frac{2.5 \times 10^5 I(s)}{s} + \frac{5}{s} \quad (3-3-2)$$

where

$$I(s) = \mathcal{L}[i] \quad (3-3-3)$$

Eq. 3-3-2 can be solved for $I(s)$ as

$$I(s) = \frac{0.04(s + 62.5)}{s^2 + 62.5s + 1.25 \times 10^5} \quad (3-3-4)$$

The denominator has the following zeros

$$s_1 \text{ and } s_2 = -31.25 \pm j353 \quad (3-3-5)$$

and Eq. 3-3-4 can be written as

$$I(s) = \frac{0.04(s + 62.5)}{(s + 31.25)^2 + (353)^2} \quad (3-3-6)$$

The inverse transform can readily be obtained by using the transform table in the Appendix. However, Eq. 3-3-6 can be divided into two parts so that inverse transforms

$$\mathcal{L}^{-1} \left[\frac{s + a}{(s + a)^2 + \omega^2} \right] = e^{-at} \cos \omega t \quad (3-3-7)$$

and

$$\mathcal{L}^{-1} \left[\frac{\omega}{(s + a)^2 + \omega^2} \right] = e^{-at} \sin \omega t \quad (3-3-8)$$

derived in Chapter 2 can be used. When this is done, $I(s)$ can be written as

$$I(s) = 0.04 \left\{ \frac{s + 31.25}{(s + 31.25)^2 + (353)^2} + \frac{31.25}{353} \left[\frac{353}{(s + 31.25)^2 + (353)^2} \right] \right\} \quad (3-3-9)$$

The inverse transform is

$$i = 0.04e^{-31.25t} [\cos 353t + 0.0885 \sin 353t]$$

which can be written as

$$i = e^{-31.25t} [40 \times 10^{-3} \cos 353t + 3.54 \times 10^{-3} \sin 353t] \quad (3-3-10)$$

The Laplace transformation method of solution introduces the initial condition into the transformed equation and no auxiliary steps are necessary to solve for the arbitrary constants.

After the equation for the current has been determined, the voltages can be found as

$$e_R = iR; e_L = L \frac{di}{dt}; \text{ and } e_C = \frac{1}{C} \int_0^t i dt + e(0) \quad (3-3-11)$$

At times it is convenient to remain in the transformed domain when another quantity is needed and to perform the inverse transform operation to find the desired quantity. If the voltage e_L is desired, it can readily be found as

$$e_L = L \frac{di}{dt} \quad (3-3-12)$$

In this case the following procedure does not yield any great saving in effort but is offered to demonstrate the method.

The transform of Eq. 3-3-12 is

$$E_L(s) = 2[sI(s) - 0.04] \quad (3-3-13)$$

When Eq. 3-3-6 is substituted for $I(s)$ in Eq. 3-3-13, the result is

$$E_L(s) = \frac{0.08s(s + 62.5)}{s^2 + 62.5s + 1.25 \times 10^5} - 0.08 \quad (3-3-14)$$

The inverse transform for the first term of Eq. 3-3-14 contains an impulse and the second term is the transform of an impulse. An inspection of the circuit shows that no impulse is to be expected in e_L , so that these two impulses must cancel each other. Therefore, when the two terms of Eq. 3-3-14 are combined the resulting equation

$$E_L(s) = -\frac{10,000}{353} \left[\frac{353}{(s + 31.25)^2 + (353)^2} \right] \quad (3-3-15)$$

is greatly simplified. The inverse transform yields

$$e_L = -28.8e^{-31.25t} \sin 353t \quad (3-3-16)$$

3-4. Kirchhoff's Current Law. Kirchhoff's current law states that the algebraic sum of currents leaving a junction is equal to zero. The law can also be expressed by stating that the sum of currents entering a junction is equal to the sum of currents leaving a junction.

In the general case, Kirchhoff's current law must be written at a number of junctions. For present purposes a simple example is chosen to demonstrate the use of Laplace transforms on another simple circuit problem.

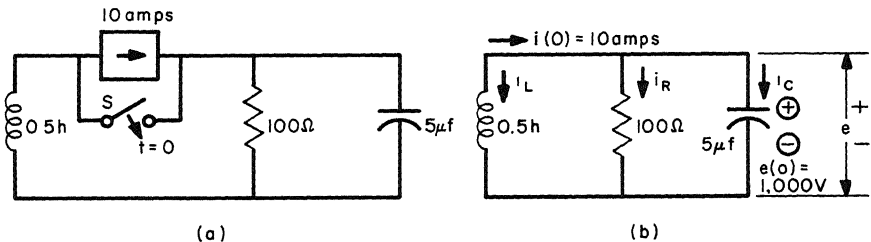


Fig. 3-4-1.

Example 3-4-1. The circuit of Fig. 3-4-1a is in steady state with the switch S open. S is then closed at $t = 0$.

Before $t = 0$, the current source supplies 10 amperes to the circuit and the path of this current is down through the 100 ohm resistor and up through the 0.5 henry inductor. The voltage across the $5\mu f$ capacitor is equal to the voltage across the resistor which is equal to 1,000 volts.

After S is closed, the resulting circuit and the initial current and voltages are shown in Fig. 3-4-1b. After the arbitrary $+$ and $-$ polarity reference for the voltage e is chosen, it is convenient (but not necessary) to select the reference direction for

each current so that it is directed into the + terminal for its respective element. An equation based on Kirchhoff's current law can now be written as

$$0 = i_L + i_R + i_c$$

which can be rewritten as

$$0 = \frac{1}{0.5} \int_0^t e \, dt - 10 + \frac{e}{100} + 5 \times 10^{-6} \frac{de}{dt} \quad (3-4-1)$$

The transformed equation is

$$0 = \frac{2E(s)}{s} - \frac{10}{s} + 0.01E(s) + 5 \times 10^{-6}[sE(s) - 1000] \quad (3-4-2)$$

where $E(s) = \mathcal{L}[e]$. Eq. 3-4-2 can be solved for $E(s)$ as

$$E(s) = \frac{1000s + 2 \times 10^6}{s^2 + 2000s + 4.0 \times 10^5} \quad (3-4-3)$$

The denominator has the following zeros:

$$s_1 = -1775 \text{ and } s_2 = -225 \quad (3-4-4)$$

Eq. 3-4-4 can be written as

$$E(s) = \frac{1000s + 2 \times 10^6}{(s + 1775)(s + 225)} \quad (3-4-5)$$

and can be expanded as

$$E(s) = -\frac{145}{(s + 1775)} + \frac{1145}{(s + 225)} \quad (3-4-6)$$

The inverse transform is

$$e = -145e^{-1775t} + 1145e^{-225t} \quad (3-4-7)$$

To continue the example, the current in the inductor is determined. This current is related to the voltage e by the equation

$$i_L = \frac{1}{L} \int_0^t e \, dt + i(0) = 2 \int_0^t e \, dt - 10 \quad (3-4-8)$$

The transform of i_L is

$$\begin{aligned} I_L(s) &= \frac{2E(s)}{s} - \frac{10}{s} \\ &= \frac{2000s + 4 \times 10^6}{s(s + 1775)(s + 225)} - \frac{10}{s} \end{aligned} \quad (3-4-9)$$

An inspection of the circuit reveals that i_L has the value zero at $t = \infty$. Therefore, when the two terms of Eq. 3-4-9 are combined, the resulting function does not have a pole at the origin.

$$I_L(s) = -\frac{10s + 18,000}{(s + 1775)(s + 225)} \quad (3-4-10)$$

This is expanded as

$$I_L(s) = \frac{0.1615}{s + 1775} - \frac{10.1615}{s + 225} \quad (3-4-11)$$

The inverse transform is

$$i_L = 0.1615e^{-1775t} - 10.1615e^{-225t} \quad (3-4-12)$$

3-5. Operational Approach to Circuits. It is frequently convenient to write the transformed equation (or equations) directly without first writing the differential equation and then performing the transform operation. As a start in developing this technique, the operational impedance of a circuit is defined as the ratio of the transform of the voltage to the transform of the current under the assumption that the circuit is initially "at rest"; that is, all initial conditions are zero.

This operational impedance concept is demonstrated by using the inductor shown in Fig. 3-5-1a as an example. The circuit equation as a differential equation is

$$e = L \frac{di}{dt} \quad (3-5-1)$$

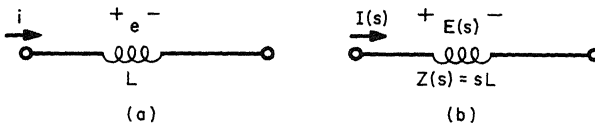


Fig. 3-5-1.

and the transform, assuming no stored energy, is

$$E(s) = sLI(s) \quad (3-5-2)$$

Therefore, the operational impedance for the inductor is

$$Z(s) = \frac{E(s)}{I(s)} = sL \quad (3-5-3)$$

Fig. 3-5-1b indicates the transformed quantities.

The operational impedances of the R and C elements are respectively

$$R \quad \text{and} \quad \frac{1}{sC} \quad (3-5-4)$$

The operational admittance of a circuit is defined as the ratio of the transform of the current to the transform of the voltage under the assumption that the circuit is initially at rest. The operational admittances of the R , L , and C elements respectively are

$$\frac{1}{R}, \frac{1}{sL}, \quad \text{and} \quad sC \quad (3-5-5)$$

IMPEDANCES IN SERIES. If a number of operational impedances are connected in series, the total impedance is the sum of the individual impedances. This is demonstrated by the following development. When Kirchhoff's voltage law is applied to the circuit of Fig. 3-5-2, the following equation can be written:

$$e = e_1 + e_2 + e_3 \quad (3-5-6)$$

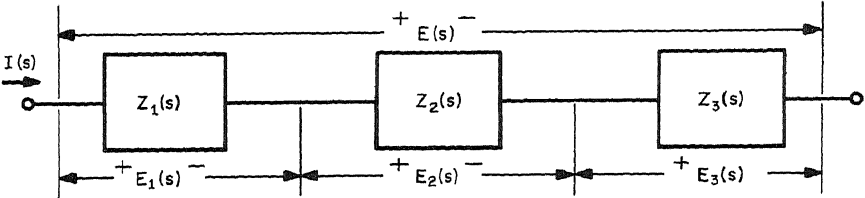


Fig. 3-5-2.

The transform of this equation is

$$E(s) = E_1(s) + E_2(s) + E_3(s) \quad (3-5-7)$$

which can be rewritten as

$$I(s)Z(s) = I(s)Z_1(s) + I(s)Z_2(s) + I(s)Z_3(s) \quad (3-5-8)$$

where $Z(s)$ is the equivalent impedance of the entire circuit. When the $I(s)$ is canceled on both sides of the equation, the result is

$$Z(s) = Z_1(s) + Z_2(s) + Z_3(s) \quad (3-5-9)$$

The manner in which the voltage divides among impedances in series is now developed. The voltage $E_2(s)$ is chosen arbitrarily to indicate the method. The current $I(s)$ is related to the total voltage and total impedance by

$$I(s) = \frac{E(s)}{Z(s)} \quad (3-5-10)$$

The current is also related to $E_2(s)$ and $Z_2(s)$ by

$$I(s) = \frac{E_2(s)}{Z_2(s)} \quad (3-5-11)$$

When $I(s)$ is eliminated between the last two equations, the result is

$$\frac{E_2(s)}{Z_2(s)} = \frac{E(s)}{Z(s)} \quad (3-5-12)$$

which can be rewritten as

$$E_2(s) = \frac{Z_2(s)}{Z(s)} E(s) \quad (3-5-13)$$

Impedances in series act as a voltage divider and the ratio of $Z_2(s)$ to $Z(s)$ determines $E_2(s)$, the portion of $E(s)$ appearing across $Z_2(s)$.

ADMITTANCES IN PARALLEL. If a number of operational admittances are connected in parallel, the total admittance is the sum of the individual admittances. This is demonstrated by the following development. When Kirchhoff's current law is applied to the circuit of Fig. 3-5-3, the following equation can be written

$$i = i_1 + i_2 + i_3 \quad (3-5-14)$$

The transform of this equation is

$$I(s) = I_1(s) + I_2(s) + I_3(s) \quad (3-5-15)$$

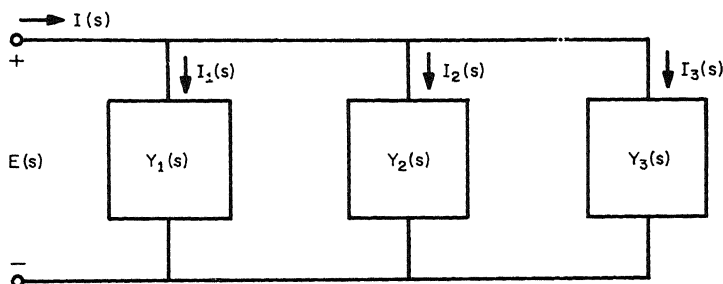


Fig. 3-5-3.

which can be written as

$$E(s)Y(s) = E(s)Y_1(s) + E(s)Y_2(s) + E(s)Y_3(s) \quad (3-5-16)$$

where $Y(s)$ is the equivalent admittance of the entire circuit. When the $E(s)$ is canceled on both sides of the equation, the result is

$$Y(s) = Y_1(s) + Y_2(s) + Y_3(s) \quad (3-5-17)$$

The manner in which the current divides among admittances in parallel is now developed. The current $I_2(s)$ is chosen arbitrarily to indicate the method. The voltage $E(s)$ is related to the total current and total admittance by

$$E(s) = \frac{I(s)}{Y(s)} \quad (3-5-18)$$

The voltage is also related to $I_2(s)$ and $Y_2(s)$ by

$$E(s) = \frac{I_2(s)}{Y_2(s)} \quad (3-5-19)$$

when $E(s)$ is eliminated between the last two equations, the result is

$$\frac{I_2(s)}{Y_2(s)} = \frac{I(s)}{Y(s)} \quad (3-5-20)$$

which can be rewritten as

$$I_2(s) = \frac{Y_2(s)}{Y(s)} I(s) \quad (3-5-21)$$

Admittances in parallel act as a current divider and the ratio of $Y_2(s)$ to $Y(s)$ determines $I_2(s)$, the portion of $I(s)$ existing in $Y_2(s)$.

The following example makes use of these concepts.

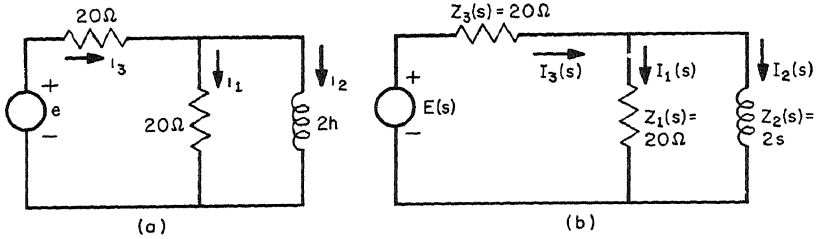


Fig. 3-5-4.

Example 3-5-1. The circuit shown in Fig. 3-5-4a is at rest when the voltage e is applied to the circuit at $t = 0$. The circuit of Fig. 3-5-4b indicates the transformed quantities. The impedance of the parallel elements are

$$Z_p(s) = \frac{1}{\frac{1}{Z_1(s)} + \frac{1}{Z_2(s)}} = \frac{Z_1(s)Z_2(s)}{Z_1(s) + Z_2(s)} = \frac{40s}{2s + 20} \quad (3-5-22)$$

The total impedance into which the voltage source delivers current is

$$Z(s) = Z_3(s) + Z_p(s) = 20 + \frac{40s}{2s + 20} = \frac{80s + 400}{2s + 20} \quad (3-5-23)$$

The current $I_3(s)$ is

$$I_3(s) = \frac{E(s)}{Z(s)} = \frac{(2s + 20)E(s)}{(80s + 400)} \quad (3-5-24)$$

The current $I_3(s)$ divides into $I_1(s)$ and $I_2(s)$. $I_2(s)$ is determined in order to demonstrate the method.

$$\begin{aligned} I_2(s) &= \frac{Y_2(s)}{Y_1(s) + Y_2(s)} I_3(s) = \frac{Z_1(s)}{Z_1(s) + Z_2(s)} I_3(s) \\ &= \frac{20(2s + 20)}{(2s + 20)(80s + 400)} E(s) = \frac{E(s)}{4(s + 5)} \end{aligned} \quad (3-5-25)$$

The voltage $E_p(s)$ across the elements in parallel can be found by using the voltage divider concept as

$$E_p(s) = \frac{Z_p(s)}{Z(s)} E(s) = \frac{40s/(2s + 20)}{(80s + 400)/(2s + 20)} E(s) = \frac{s}{2(s + 5)} E(s) \quad (3-5-26)$$

This voltage can alternately be found as

$$E_p(s) = I_2(s)Z_2(s) = \frac{E(s)2s}{4(s + 5)} = \frac{s}{2(s + 5)} E(s) \quad (3-5-27)$$

As soon as the driving function e is specified, the time functions corresponding to each of these quantities can be determined by finding the inverse transform. This is done for three different driving functions.

Unit-Step Driving Function. The first of these driving functions is

$$e = u_{-1}(t) \quad (3-5-28)$$

The corresponding $E(s)$ is

$$E(s) = \frac{1}{s} \quad (3-5-29)$$

With this $E(s)$ substituted into Eq. 3-5-24 for $I_3(s)$, the equation becomes

$$I_3(s) = \frac{s + 10}{40s(s + 5)} = \frac{0.05}{s} - \frac{0.025}{s + 5} \quad (3-5-30)$$

The time function is

$$i_3 = 0.05 - 0.025e^{-5t} \quad (3-5-31)$$

With the $E(s)$ of Eq. 3-5-29 substituted into Eq. 3-5-25 for $I_2(s)$, the equation can be written as

$$I_2(s) = \frac{1}{4s(s + 5)} = \frac{0.05}{s} - \frac{0.05}{s + 5} \quad (3-5-32)$$

The time function is

$$i_2 = 0.05(1 - e^{-5t}) \quad (3-5-33)$$

With the $E(s)$ of 3-5-29 substituted into Eq. 3-5-26 for $E_p(s)$, the equation is

$$E_p(s) = \frac{1}{2(s + 5)} \quad (3-5-34)$$

The time function is

$$e_p = 0.5e^{-5t} \quad (3-5-35)$$

Unit-Ramp Driving Function. The second of these driving functions is

$$e = u_{-2}(t) \quad (3-5-36)$$

The corresponding $E(s)$ is

$$E(s) = \frac{1}{s^2} \quad (3-5-37)$$

With this $E(s)$ substituted into Eq. 3-5-24, the equation for $I_3(s)$ becomes

$$I_3(s) = \frac{s + 10}{40s^2(s + 5)} = \frac{0.05}{s^2} - \frac{0.005}{s} + \frac{0.005}{s + 5} \quad (3-5-38)$$

The time function is

$$i_3 = 0.05t - 0.005 + 0.005e^{-5t} \quad (3-5-39)$$

Another interesting property resulting from linearity can be commented upon here. When a system can be described by a set of linear differential equations, the driving function and the response function are interrelated. For example, if the

response to a certain driving function is known, the response to the integral (or derivative) of the driving function can be found by integrating (or differentiating) the original response. In this example, the response function to a unit-step is given in Eq. 3-5-31, repeated for convenience as

$$i_3 = 0.05 - 0.025e^{-5t} \quad (3-5-40)$$

The response to a unit-ramp function can be found by integrating this response as

$$\begin{aligned} i_3 &= \int_0^t 0.05 \, dt - \int_0^t 0.025e^{-5t} \, dt \\ &= 0.05t \Big|_0^t + 0.005e^{-5t} \Big|_0^t \\ &= 0.05t - 0.005 + 0.005e^{-5t} \end{aligned} \quad (3-5-41)$$

This result checks Eq. 3-5-39 as found previously.

With the $E(s)$ of Eq. 3-5-37 substituted into Eq. 3-5-25, the equation for $I_2(s)$ becomes

$$\begin{aligned} I_2(s) &= \frac{1}{4s^2(s+5)} \\ &= \frac{0.05}{s^2} - \frac{0.01}{s} + \frac{0.01}{s+5} \end{aligned} \quad (3-5-42)$$

The time function is

$$i_2 = 0.05t - 0.01 + 0.01e^{-5t} \quad (3-5-43)$$

This equation is also determined by integrating Eq. 3-5-33 as

$$\begin{aligned} i_2 &= \int_0^t 0.05 \, dt - \int_0^t 0.05e^{-5t} \, dt \\ &= 0.05t \Big|_0^t + 0.01e^{-5t} \Big|_0^t \\ &= 0.05t - 0.01 + 0.01e^{-5t} \end{aligned} \quad (3-5-44)$$

The equation for e_p is found by solving Prob. 3-9.

Unit-Impulse Driving Function. The third of these driving functions is

$$e = u_o(t) \quad (3-5-45)$$

The corresponding $E(s)$ is

$$E(s) = 1 \quad (3-5-46)$$

With this $E(s)$ substituted into Eq. 3-5-24, the equation for $I_3(s)$ becomes

$$I_3(s) = \frac{s+10}{40s+200} \quad (3-5-47)$$

To expand this, the denominator is divided into the numerator, yielding

$$I_3(s) = 0.025 + \frac{5}{40(s+5)} \quad (3-5-48)$$

The time function is

$$i_3 = 0.025u_0(t) + 0.125e^{-5t} \quad (3-5-49)$$

The unit-impulse is the derivative of the unit-step. Therefore, Eq. 3-5-49 can also be found as the derivative of Eq. 3-5-31 which is repeated for convenience as

$$\dot{i}_3 = 0.05 - 0.025e^{-5t} \quad (3-5-50)$$

Since this equation has a discontinuity of 0.025 at $t = 0$, its derivative contains an impulse with a value of 0.025 occurring at $t = 0$. Once the impulse is determined, the derivative is taken in the usual manner yielding

$$i_3 = 0.025u_0(t) + 0.125e^{-5t} \quad (3-5-51)$$

With the $E(s)$ of Eq. 3-5-46 substituted into Eq. 3-5-25, the equation for $I_2(s)$ becomes

$$I_2(s) = \frac{1}{4(s + 5)} \quad (3-5-52)$$

The time function is

$$i_2 = 0.25e^{-5t} \quad (3-5-53)$$

This equation can also be found as the derivative of Eq. 3-5-33 which is repeated for convenience as

$$\dot{i}_2 = 0.05(1 - e^{-5t}) \quad (3-5-54)$$

This equation does not have a discontinuity at $t = 0$ and the derivative is taken in the conventional manner as

$$i_2 = 0.25e^{-5t} \quad (3-5-55)$$

The equation for e_p is found by solving Prob. 3.10.

The practice of differentiating (or integrating) the driving function and the response function can be continued indefinitely. Furthermore, the concept of convolution makes it possible to find the response of a linear physical system to any arbitrary input as soon as its response to a unit-step input is known. The convolution integral is briefly introduced in Chapter 2 and is used in more detail in Chapter 10.

3-6. Particular Solutions for Sinusoidal Driving Functions.

The sinusoid as a driving function occurs so frequently that a body of specialized theory and technique has been developed in electrical engineering for finding the particular solution. This body of theory and technique comprises alternating-current (a-c) circuit theory. This theory is much broader in scope than the term a-c circuit implies, since it can be used whenever a physical system can be described by a linear differential equation with constant coefficients.

As this book assumes the reader has a knowledge of fundamental a-c circuit theory, the following discussion is intended only to bring out the

relationships between this theory and the operational approach presented earlier in this chapter.

All the operations of a-c circuit theory can be performed when trigonometric functions are needed to represent the instantaneous values involved. However, some of these operations are rather unwieldy, and in order to simplify the manipulation of trigonometric functions the concept of phasors is helpful. A phasor is represented by a directed line segment that rotates in the counterclockwise direction at a constant angular velocity; the projection of the magnitude of the phasor on a horizontal line (or on a vertical line) is equal to the instantaneous value of the sinusoidal function being represented.

This representation of instantaneous values by a phasor is actually a transformation and the term *phasor transform* is used in this book to signify this change in domain.

A convenient mathematical method of representing a phasor (current is used as an example) is

$$i = \text{Re} [I_m e^{j(\omega t + \varphi)}] \quad (3-6-1)$$

where I_m represents the maximum value of the current and the symbol Re stands for the real part of the function inside the bracket. This function inside the bracket represents the phasor and the process of taking the real part is equivalent to projecting the phasor on a horizontal line.

When the current is differentiated, a $j\omega$ factor multiplies the term inside the bracket representing the phasor, as in

$$\frac{di}{dt} = \text{Re}[j\omega I_m e^{j(\omega t + \varphi)}] \quad (3-6-2)$$

If the current is integrated, a $j\omega$ factor divides the term inside the bracket representing the phasor, as in

$$\int i \, dt = \text{Re} \left[\frac{I_m e^{j(\omega t + \varphi)}}{j\omega} \right] \quad (3-6-3)$$

These equations should be compared with similar expressions obtained from the operational methods in which initial conditions are neglected. The Laplace transform of a current is

$$\mathcal{L}[i] = I(s) \quad (3-6-4)$$

The transforms for the derivative and integral of the current are

$$\mathcal{L}\left[\frac{di}{dt}\right] = sI(s) \quad (3-6-5)$$

$$\mathcal{L}\left[\int i \, dt\right] = \frac{I(s)}{s} \quad (3-6-6)$$

A comparison of Eq. 3-6-5 with Eq. 3-6-2 and Eq. 3-6-6 with Eq. 3-6-3 shows that s appears in the Laplace transforms in exactly the same location as $j\omega$ appears in the phasor transforms. Therefore, to obtain the phasor transform of $f(t)$ formally from the Laplace transform of a differential equation, where all initial conditions are quiescent, the quantity s is replaced by $j\omega$ and the Laplace transform $\mathcal{L}[f(t)] = F(s)$ is replaced by the phasor transform $F(j\omega)$. The s in $F(s)$ signifies that the function is a Laplace transform whereas the $j\omega$ in $F(j\omega)$ signifies a phasor transform.

The substitution $s = j\omega$ can always be made. If the quantity represented by a phasor is differentiated twice, a $(j\omega)^2$ factor appears; similarly the Laplace transform has an s^2 factor. If the quantity represented by a phasor is integrated twice a $1/(j\omega)^2$ factor appears; similarly the Laplace transform has a $1/s^2$ factor. As this line of reasoning can be extended to a derivative or integral of any order, the correspondance of $j\omega$ and s is complete.

In Art. 3-5, the operational impedance is defined as the ratio of the Laplace transform whereas the $j\omega$ in $F(j\omega)$ signifies a phasor transform. thus

$$Z(s) = \frac{E(s)}{I(s)} \quad (3-6-7)$$

The phasor impedance can be defined in an analogous manner as

$$Z(j\omega) = \frac{E(j\omega)}{I(j\omega)} \quad (3-6-8)$$

where $E(j\omega)$ and $I(j\omega)$ are the phasor transforms of the voltage and current.

The operational and phasor impedances and admittances for each of the circuit elements are shown in Table 3-6-1.

TABLE 3-6-1

Circuit element	Operational impedance	Phasor impedance	Operational admittance	Phasor admittance
Resistor	R	R	$1/R$	$1/R$
Inductor	sL	$j\omega L$	$1/sL$	$1/j\omega L$
Capacitor	$1/sC$	$1/j\omega C$	sC	$j\omega C$

All the concepts developed for operational quantities also hold for phasor quantities. For example, it is shown in Art. 3-5 that if a group of operational impedances is connected in series, the total impedance is the sum of the individual impedances. In the case of phasor impedances, this equation becomes

$$Z(j\omega) = Z_1(j\omega) + Z_2(j\omega) + Z_3(j\omega) + \cdots \quad (3-6-9)$$

Similarly if a group of admittances are connected in parallel, the total phasor admittance is

$$Y(j\omega) = Y_1(j\omega) + Y_2(j\omega) + Y_3(j\omega) + \cdots \quad (3-6-10)$$

The circuit of Example 3-5-1 is used to demonstrate these concepts.

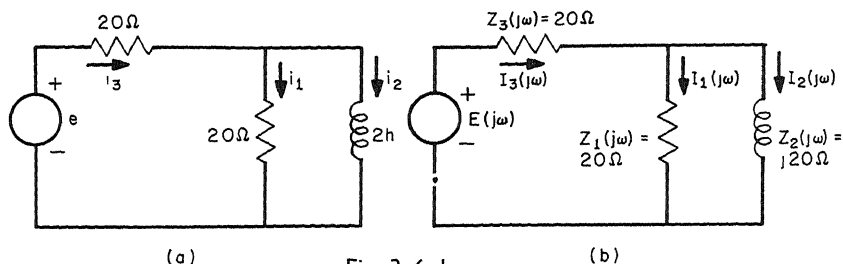


Fig. 3-6-1.

Example 3-6-1. The circuit of Fig. 3-5-4a is redrawn as Fig. 3-6-1a. The voltage

$$e = \cos 10t \quad (3-6-11)$$

is applied to the circuit and is left until steady-state conditions are reached. The current i_2 in steady state is the desired solution. Eq. 3-5-25 for $I_2(s)$ repeated here for convenience is the starting point:

$$I_2(s) = \frac{E(s)}{4(s + 5)} \quad (3-6-12)$$

The voltage e can be expressed in phasor form as

$$E(j\omega) = 1 \angle 0^\circ \quad (3-6-13)$$

The peak value of voltage is used, and the phase angle is chosen, with the assumption that the phasor is projected on a horizontal line.

Eq. 3-6-12 can be replaced by the corresponding phasor equation by replacing $E(s)$ by $E(j\omega)$ as given in Eq. 3-6-13, $I_2(s)$ by $I_2(j\omega)$ and s by $j\omega$. Upon substitution the equation becomes

$$I_2(j\omega) = \frac{1 \angle 0^\circ}{4(j10 + 5)} = 0.0223 \angle -63.5^\circ \quad (3-6-14)$$

The projection of this phasor on the horizontal line yields the time function

$$i_{2ss} = 0.0223 \cos(10t - 63.5^\circ) \quad (3-6-15)$$

The substitution of $s = j\omega$ in order to proceed from the Laplace transform of a quantity to the phasor transform is used in a number of places throughout the book.

3-7. The Complete Solution for Sinusoidal Driving Functions.

The previous article discusses the manner in which the particular solution for a sinusoidal driving function can be found from the corresponding Laplace transform. If the complete solution is desired, both the particular and complementary components of the solution are found automatically if Laplace transform methods are used.

Example 3-7-1. The circuit of Fig. 3-6-1 is again used as an example with the driving function

$$e = [\cos 10t]u_{-1}(t) \quad (3-7-1)$$

and the current i_2 is the desired response. The cosine function multiplied by $u_{-1}(t)$ signifies that the cosine function is applied to the circuit at $t = 0$.

Eq. 3-5-25 can again be used and is repeated as

$$I_2(s) = \frac{E(s)}{4(s + 5)} \quad (3-7-2)$$

In this case the Laplace transform of e is given by

$$E(s) = \frac{s}{s^2 + (10)^2} \quad (3-7-3)$$

Upon substitution Eq. 3-7-2 becomes

$$I_2(s) = \frac{s}{4[s^2 + (10)^2](s + 5)} \quad (3-7-4)$$

$I_2(s)$ is expanded as

$$\begin{aligned} I_2(s) &= \frac{s}{4(s - j10)(s + j10)(s + 5)} \\ &= \frac{0.0223 \angle -63.5^\circ}{2(s - j10)} + \frac{0.0223 \angle +63.5^\circ}{2(s + j10)} - \frac{0.01}{s + 5} \end{aligned} \quad (3-7-5)$$

The inverse transform is

$$i_2 = \frac{0.0223 \angle -63.5^\circ}{2} e^{+j10t} + \frac{0.0223 \angle +63.5^\circ}{2} e^{-j10t} - 0.01e^{-5t} \quad (3-7-6)$$

The first two terms can be combined into a cosine function as follows:

$$i_2 = 0.0223 \left[\frac{e^{+j(10t - 63.5/57.3)} + e^{-j(10t - 63.5/57.3)}}{2} \right] - 0.01e^{-5t} \quad (3-7-7)$$

$$i_2 = 0.0223 \cos(10t - 63.5^\circ) - 0.01e^{-5t} \quad (3-7-8)$$

The first term of this solution is the particular or steady-state component and checks with Eq. 3-6-15, whereas the second term is the complementary or transient component.

The use of Laplace transforms offers a very convenient method of handling many different problems. If the steady-state sinusoidal response is desired, the transforms of the driving function and the response function are replaced by phasor transforms and s is replaced by $j\omega$ throughout the remainder of the expression. If the complete solution is desired, the driving function is transformed and the response time function can be found as the inverse transform.

3-8. Transfer Functions. In order to encompass the wide variety of physical systems that are encountered in control systems, the operational

approach to circuit theory can be broadened to include all physical systems. As a start in this direction the term transfer function is defined as the ratio of the Laplace transform of the output to the Laplace transform of the input of the system being considered. The differential equation is transformed assuming that there is no energy stored in the system before $t = 0$: i. e., all initial conditions are zero. The transfer function is labeled $G(s)$ and is given by

$$G(s) = \frac{O(s)}{I(s)} \quad (3-8-1)$$

where $G(s)$ is the transfer function,

$O(s)$ is the Laplace transform of the output,

$I(s)$ is the Laplace transform of the input.

The definition of transfer function is completely general in the sense that the input and output can be any combination of voltage, current, displacement, pressure temperature, torque, or any other physical quantity. Only transfer functions for electric circuits are considered in this chapter; the transfer functions for other systems are discussed in following chapters.

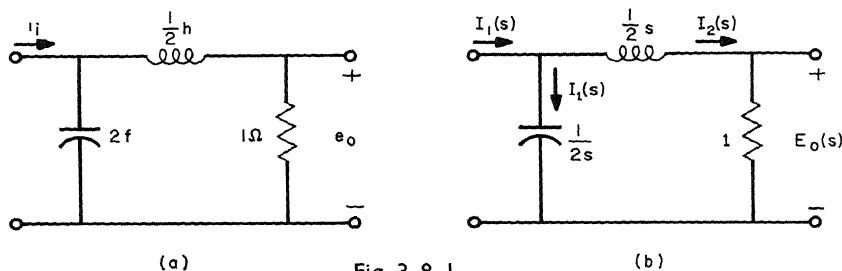


Fig. 3-8-1.

Example 3-8-1. The circuit of Fig. 3-8-1a is driven by a current i_i and the output is a voltage e_o . The desired transfer function is the ratio of the Laplace transform of e_o to the Laplace transform of i_i . The circuit of Fig. 3-8-1b shows the transformed quantities. The current $I_i(s)$ divides into $I_1(s)$ and $I_2(s)$. $I_2(s)$ can be found as

$$I_2(s) = \frac{\frac{1}{2s} I_i(s)}{\frac{1}{2s} + \frac{1}{2}s + 1} = \frac{I_i(s)}{(s+1)^2} \quad (3-8-2)$$

The voltage $E_o(s)$ is

$$E_o(s) = I_2(s)1 = \frac{I_i(s)}{(s+1)^2} \quad (3-8-3)$$

Finally the ratio of output to input is

$$G(s) = \frac{E_o(s)}{I_i(s)} = \frac{1}{(s+1)^2} \quad (3-8-4)$$

It is hoped that use of a value of capacitance equal to 2 farads does not disturb the reader as it is well known that this represents a large capacitor. With the ability to scale from one set of element values to another set, the

choice of values in a numerical example is immaterial. The subject of scaling is discussed in detail in Chapter 14.

In this example the transfer function is a ratio of voltage to current and is therefore termed a transfer impedance, since the current is supplied at one set of terminals and the voltage is obtained at another set of terminals. If the voltage and current are both associated with the same set of terminals, the impedance is called a driving point impedance.

As soon as the transfer function of a system is found, a block diagram can be drawn which contains all the information about the system as far as mathematical analysis is concerned. The block diagram for this

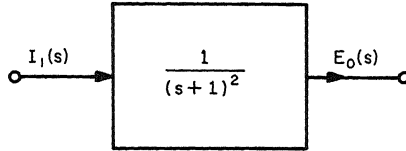


Fig. 3-8-2.

example is shown in Fig. 3-8-2. The arrows on the diagram indicate the direction of signal flow. This block diagram does not carry any information concerning the types of elements or the manner in which they are connected in the circuit, but it does contain all the information about the input, the output, and the behavior of the circuit.

Eq. 3-8-1 is the defining equation for transfer functions and can be solved as

$$O(s) = G(s)I(s)$$

If the input is given as a time function, $I(s)$ can be found as the Laplace transform of this function. With $G(s)$ known, the product of $G(s)$ is the transform of the output. The output as a time function can be found as the inverse transform of this product. This technique is used in Art. 3-5 to solve for i_3 , i_2 , and e_p for various driving functions as given, even though the term transfer function is not used at that time.

However, this method can be used only when the system does not contain stored energy at $t = 0$. This limitation is removed in Arts. 3-10 and 3-11.

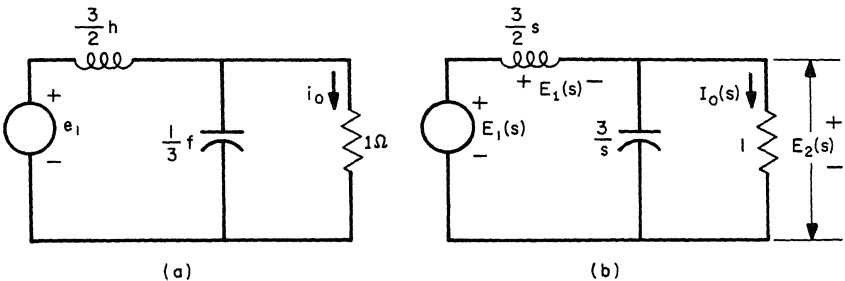


Fig. 3-8-3.

Example 3-8-2. The circuit of Fig. 3-8-3a is driven by a voltage and the output is a current. This circuit of Fig. 3-8-3b shows the transformed quantities. The voltage $E_1(s)$ divides into $E_1(s)$ and $E_2(s)$. $E_2(s)$ can be found as

$$E_2(s) = \frac{\frac{2}{3s} E_1(s)}{\frac{2}{3s} + \frac{s}{3} + 1} = \frac{2E_1(s)}{(s+1)(s+2)} \quad (3-8-5)$$

The current $I_o(s)$ is

$$I_o(s) = E_2(s)1 = \frac{2E_1(s)}{(s+1)(s+2)} \quad (3-8-6)$$

Finally, the ratio of output to input is

$$G(s) = \frac{I_o(s)}{E_1(s)} = \frac{2}{(s+1)(s+2)} \quad (3-8-7)$$

The block diagram for this example is shown in Fig. 3-8-4.

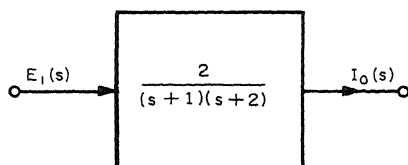


Fig. 3-8-4.

In this example, the transfer function is a ratio of current to voltage and is therefore termed a transfer admittance since the voltage is supplied to one part of the circuit and the current exists in another part of the circuit. If the voltage and current are both associated with the same set of terminals, the admittance is called a driving point admittance.

When a driving point admittance is “flipped over,” it becomes a driving point impedance. This is true because the current can be considered as the driving function and the voltage as the response, and vice versa, when only two terminals are involved. However, a similar relationship does not exist for transfer admittance and transfer impedance. To clarify this concept, an extreme example is used. If a million volts are placed on a transmission line in New York and the current out of the transmission line in California is one microampere, it does not follow that one microampere can be supplied to the transmission line in California with the result that a million volts will be produced in New York. If the driving function and response function are at two different points in the physical system, the distinction between them must be made. This distinction can always be maintained by remembering that a transfer function is a ratio of output to input.

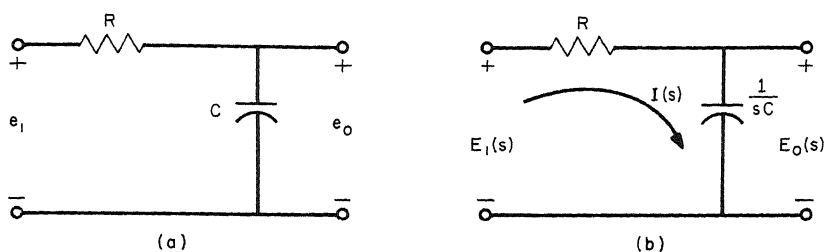


Fig. 3-8-5.

Example 3-8-3. The circuit of Fig. 3-8-5a is driven by a voltage and the output is a voltage. The circuit of Fig. 3-8-5b shows the transformed quantities. The current $I(s)$ is

$$I(s) = \frac{E_i(s)}{R + \frac{1}{Cs}} \quad (3-8-8)$$

and the voltage $E_o(s)$ is

$$E_o(s) = \frac{1}{Cs} I(s) = \frac{E_i(s)}{RCs + 1} \quad (3-8-9)$$

The ratio of output to input is

$$G(s) = \frac{E_o(s)}{E_i(s)} = \frac{1}{RCs + 1} \quad (3-8-10)$$

This transfer function is a ratio of the transform of the output voltage to the transform of the input voltage and the term transfer functions is retained.

The block diagram for this example is shown in Fig. 3-8-6.

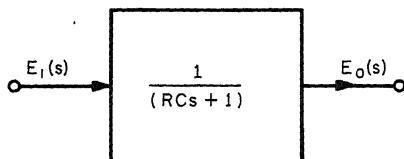


Fig. 3-8-6.

This block diagram makes the point so clearly that an interesting side issue may be explored. A given physical system has only one transfer function relating two specific variables, but the reverse of this is not true. A given transfer function can represent an infinite number of systems. In this particular example, an infinite number of circuits can have the same transfer as long as the product RC is held constant. In addition, this transfer function can represent an R - L circuit as shown in the next example. Furthermore, with the proper change of input and output variables the transfer function can represent certain mechanical, hydraulic, or many other systems.

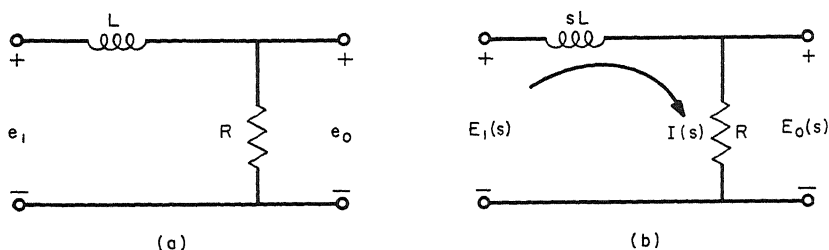


Fig. 3-8-7.

Example 3-8-4. The circuit of Fig. 3-8-7a is driven by a voltage and the output is a voltage. The circuit of Fig. 3-8-7b shows the transformed quantities. The current $I(s)$ is

$$I(s) = \frac{E_i(s)}{Ls + R} \quad (3-8-11)$$

and the voltage $E_o(s)$ is

$$E_o(s) = RI(s) = \frac{E_i(s)}{\frac{L}{R}s + 1} \quad (3-8-12)$$

The ratio of output to input is

$$G(s) = \frac{E_o(s)}{E_i(s)} = \frac{1}{\frac{L}{R}s + 1} \quad (3-8-13)$$

The block diagram for this example is shown in Fig. 3-8-8.

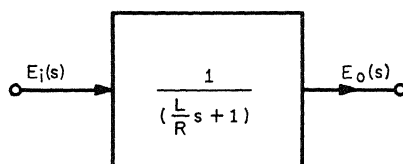


Fig. 3-8-8.

A comparison of the block diagram of Fig. 3-8-6 and Fig. 3-8-8 reveals that the two systems have identical transfer functions if the product RC in the first is equal to the ratio L/R in the second.

3-9. Transfer Functions of Network Connected in Tandem.

Two networks are said to be connected in tandem if the output of one network is the input to the next. If the second network can be connected to the first without “loading” the first network in any way, the product of the two transfer functions is the transfer function of the combined network. To demonstrate this the tandem group of networks shown in Fig.

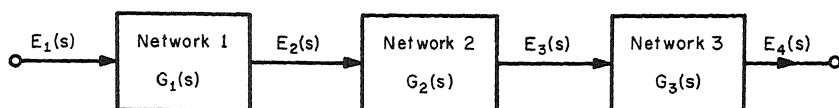


Fig. 3-9-1.

3-9-1 is considered. The transfer functions for the networks individually are

$$G_1(s) = \frac{E_2(s)}{E_1(s)} \quad (3-9-1)$$

$$G_2(s) = \frac{E_3(s)}{E_2(s)} \quad (3-9-2)$$

and

$$G_3(s) = \frac{E_4(s)}{E_3(s)} \quad (3-9-3)$$

where these transfer functions are found with each pair of the output terminals open-circuited. If the input impedance of Network 2 is infinite, the output of Network 1 is not affected by adding Network 2. Similarly the input impedance of Network 3 must be infinite in order to avoid affecting Network 2. Under these conditions the three separate transfer functions can be multiplied as

$$\frac{E_2(s)}{E_1(s)} \cdot \frac{E_3(s)}{E_2(s)} \cdot \frac{E_4(s)}{E_3(s)} = \frac{E_4(s)}{E_1(s)} \quad (3-9-4)$$

This equation can be rewritten as

$$G_1(s)G_2(s)G_3(s) = G(s) \quad (3-9-5)$$

where

$$G(s) = \frac{E_4(s)}{E_1(s)} \quad (3-9-6)$$

is the transfer function for the tandem combination.

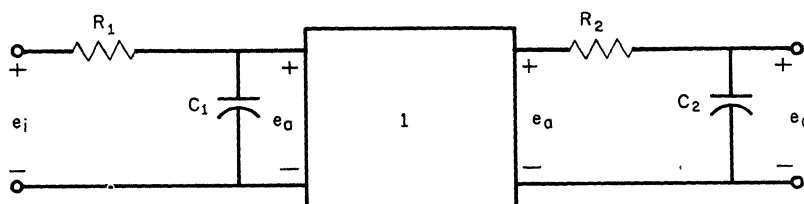


Fig. 3-9-2.

In order to examine this situation more completely, the circuit of Fig. 3-9-2 is analyzed. The box containing the number 1 represents an isolation stage so that the input of the R_2 - C_2 circuit does not load the output of the R_1 - C_1 circuit. The transfer function $G(s)$ for the entire circuit is equal to $G_1(s)G_2(s)$ as given by

$$G(s) = G_1(s)G_2(s) = \frac{1}{(R_1C_1s + 1)(R_2C_2s + 1)} \quad (3-9-7)$$

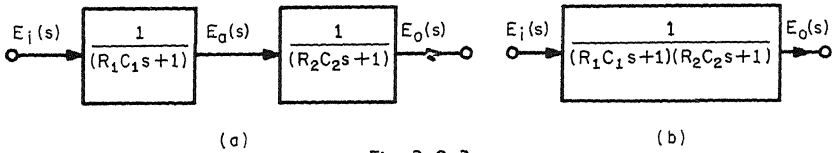


Fig. 3-9-3.

The fact that the transfer functions of the two portions of the circuit of Fig. 3-9-2 are independent is shown on a block diagram as in Fig. 3-9-3a. These two blocks can be combined into one block as shown in Fig. 3-9-3b.

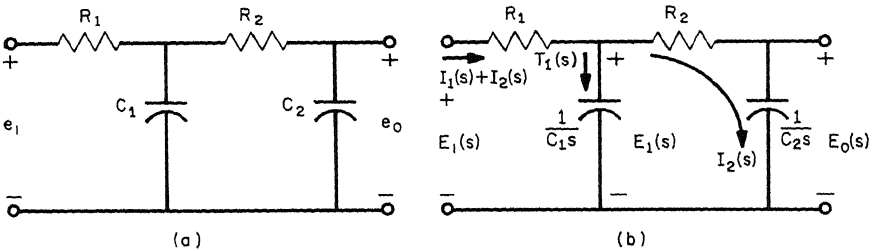


Fig. 3-9-4.

The network of Fig. 3-9-4a appears to be similar to the circuit of Fig. 3-9-2. In this case, however, the R_2 - C_2 portion of the circuit produces a loading effect on the R_1 - C_1 portion of the circuit. Therefore, the transfer function must be obtained by analyzing the entire circuit. One convenient method is to derive the circuit equations by beginning at the output and moving through the circuit to the input. In the transformed circuit shown in Fig. 3-9-4b, the voltage $E_o(s)$ produces a current $I_2(s)$ in C_2 as

$$I_2(s) = E_o(s)C_2s \quad (3-9-8)$$

The current $I_2(s)$ exists in the R_2 - C_2 portion of the circuit and the voltage $E_1(s)$ is

$$\begin{aligned} E_1(s) &= I_2(s) \left[R_2 + \frac{1}{C_2s} \right] = E_o(s)C_2s \left[R_2 + \frac{1}{C_2s} \right] \\ &= E_o(s)[R_2C_2s + 1] \end{aligned} \quad (3-9-9)$$

The voltage $E_1(s)$ produces a current $I_1(s)$ in C_1 as

$$I_1(s) = E_1(s)C_1s = E_o(s)[R_2C_2C_1s^2 + C_1s] \quad (3-9-10)$$

The total current in R_1 is

$$I_1(s) + I_2(s) = E_o(s)[R_2C_2C_1s^2 + (C_2 + C_1)s] \quad (3-9-11)$$

The input voltage is

$$\begin{aligned} E_i(s) &= [I_1(s) + I_2(s)]R_1 + E_1(s) \\ &= E_o(s)[R_2C_2R_1C_1s^2 + (R_1C_2 + R_1C_1 + R_2C_2)s + 1] \end{aligned} \quad (3-9-12)$$

The transfer function for the circuit is

$$G(s) = \frac{E_o(s)}{E_i(s)} = \frac{1}{R_2C_2R_1C_1s^2 + (R_1C_2 + R_1C_1 + R_2C_2)s + 1} \quad (3-9-13)$$

The block diagram for the circuit is shown in Fig. 3-9-5.

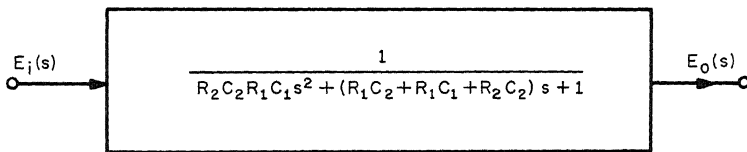


Fig. 3-9-5.

It is interesting to compare the block diagrams of Fig. 3-9-3b and Fig. 3-9-5. When the two factors of the denominator of the block diagram of Fig. 3-9-3b are multiplied, the product becomes

$$R_2C_2R_1C_1s^2 + (R_1C_1 + R_2C_2)s + 1 \quad (3-9-14)$$

The denominator of the block diagram of Fig. 3-9-5 has an extra term R_1C_2 multiplying s . This term is present in the second system because of interaction between the two portions of the circuit.

For complicated systems, the transfer function block diagram approach offers a very convenient tool. The entire system is examined and the portions thereof that are unaffected by the presence of the remainder of the system are determined. The transfer function of these independent portions can then be found. If the output of each of these portions feeds the input of the next, the overall transfer function is the product of the individual transfer functions. The situation in which the transfer functions are interconnected in a more complicated manner is studied in later chapters.

In certain cases, the loading effect of the succeeding stage can be included in the transfer function of the preceding stage. An extremely

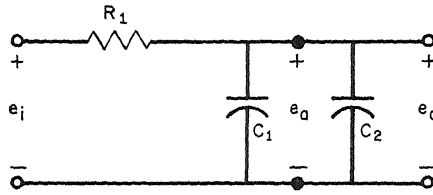


Fig. 3-9-6.

simple example of this is shown in the circuit of Fig. 3-9-6. The transfer function of the first portion of this circuit (that is, with C_2 removed) is

$$\frac{E_a(s)}{E_i(s)} = \frac{1}{R_1 C_1 s + 1} \quad (3-9-15)$$

The effect of adding the C_2 to the circuit is to increase C_1 to $C_1 + C_2$ and the transfer function of the combination is

$$\frac{E_o(s)}{E_i(s)} = \frac{1}{R_1 (C_1 + C_2) s + 1} \quad (3-9-16)$$

This idea is discussed in more detail at several points in the book.

3-10. Treating Initial Conditions. The operational approach to circuits as discussed thus far assumes that no energy is stored in the circuit at $t = 0$. To accomplish this the equations relating voltage and current for the circuit elements are transformed in Art. 3-5 assuming no initial current or voltage. The initial conditions can be added as demonstrated in the following examples. An inductor with an initial current $i(0)$ is

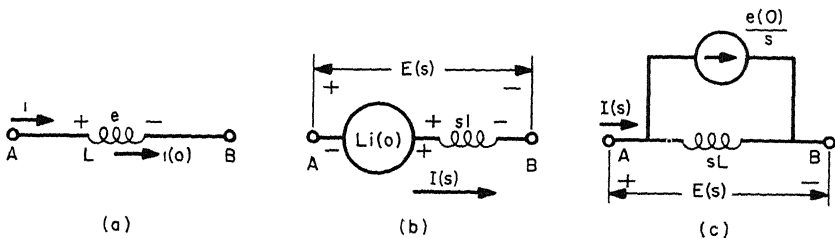


Fig. 3-10-1.

shown in Fig. 3-10-1a. The equation relating current and voltage for the inductor can be put into two forms as

$$e = L \frac{di}{dt} \quad (3-10-1)$$

and

$$i = \frac{1}{L} \int_0^t e \, dt + i(0) \quad (3-10-2)$$

The transform of Eq. 3-10-1 is given by

$$E(s) = L[sI(s) - i(0)] = sLI(s) - Li(0) \quad (3-10-3)$$

Since this equation is a voltage equation, the two terms $sLI(s)$ and $-Li(0)$ can be thought of as two voltages in series. The first term $sLI(s)$ represents a voltage across an inductor with no initial current and the second term represents an impulse of voltage. The transformed circuit of Fig. 3-10-1b is described by Eq. 3-10-3 and is therefore an equivalent circuit. At $t = 0$, the impulse of voltage establishes the current $i(0)$ in the inductor.

The transform of Eq. 3-10-2 is

$$I(s) = \frac{E(s)}{sL} + \frac{i(0)}{s} \quad (3-10-4)$$

Since this equation is a current equation, the two terms $E(s)/sL$ and $i(0)/s$ can be thought of as two currents in parallel. The first term $E(s)/sL$ represents a current that exists in an inductor with no initial current and the second term represents a constant current source. The transformed circuit of Fig. 3-10-1c is described by Eq. 3-10-4 and is therefore an equivalent circuit.

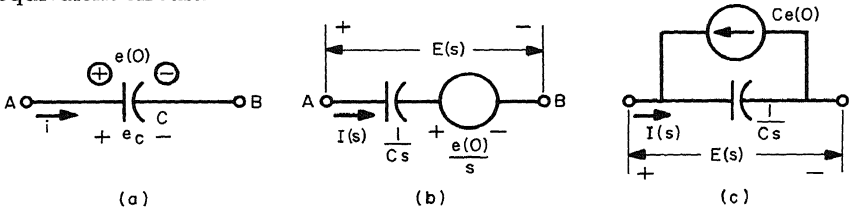


Fig. 3-10-2.

A capacitor with an initial charge (or voltage) is shown in Fig. 3-10-2a. The equation relating current and voltage for the capacitor can be put into two forms as

$$i = C \frac{de}{dt} \quad (3-10-5)$$

and

$$e = \frac{1}{C} \int_0^t i dt + e(0) \quad (3-10-6)$$

Eq. 3-10-5 and 3-10-6 are transformed as Eq. 3-10-7 and 3-10-8 respectively.

$$I(s) = C[sE(s) - e(0)] = CsE(s) - Ce(0) \quad (3-10-7)$$

$$E(s) = \frac{I(s)}{Cs} + \frac{e(0)}{s} \quad (3-10-8)$$

From these equations the equivalent circuits shown in Fig. 3-10-2b and c are obtained.

Initial conditions can be added to the operational approach to circuits by moving the initial conditions into sources. An inductor with an initial current can be replaced by an inductor with no initial current in series with an impulse type voltage source or by an inductor with no initial current in parallel with a constant current source. A capacitor with an initial charge can be replaced by an uncharged capacitor in series with a constant voltage source or by an uncharged capacitor in parallel with an impulse type current source.

It is convenient to change the initial conditions into voltage sources when solving a circuit by loop currents and into current sources when solving a circuit by nodal methods. However, it is not necessary to do so because the two types of sources can be mixed.

3-11. Vacuum Tube Circuit Transfer Functions. An extension of circuit analysis and transfer function techniques to tube circuits can be made by means of "incremental" equivalent circuits. Such circuits are valid for increments or variations from normal operating values of system variables. The following development assumes a knowledge of basic

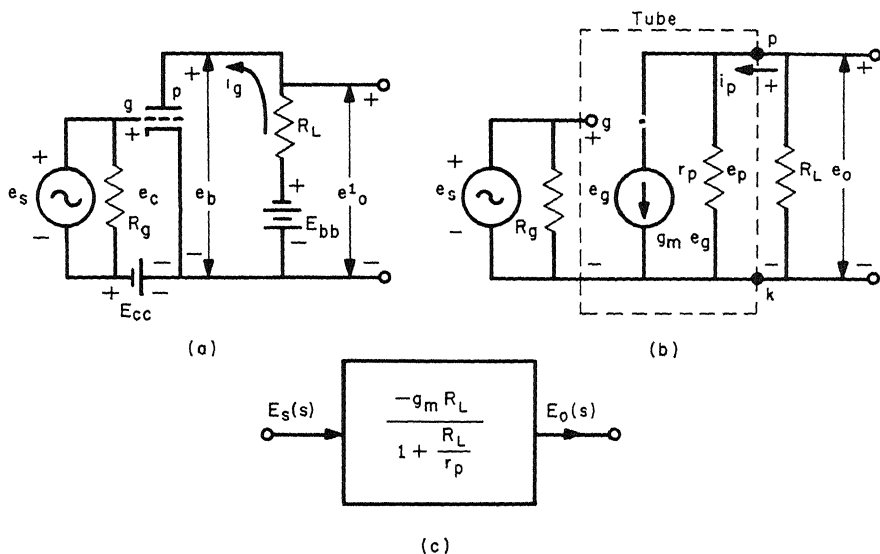


Fig. 3-11-1.

tube circuit operation. In the simple triode amplifier of Fig. 3-11-1a the total instantaneous plate current i_b is some function of the total instantaneous grid and plate voltages.

$$i_b = f_1(e_b, e_g) \quad (3-11-1)$$

Generally only the signals which are superimposed on d-c operating voltages and currents need be studied. A convenient approach is to study

increments in the variables in Eq. 3-11-1. The total differential of i_b is

$$di_b = \frac{\partial i_b}{\partial e_b} de_b + \frac{\partial i_b}{\partial e_c} de_c \quad (3-11-2)$$

For small variations in i_b , e_b , and e_c about an operating point (I_{bo} , E_{bo} , E_{co}) the differentials can be approximated by increments and the partial derivatives by their values at the operating point as follows:

$$\Delta i_b \approx \frac{1}{r_p} \Delta e_b + g_m \Delta e_c \quad (3-11-3)$$

$$i_p \approx \frac{1}{r_p} e_p + g_m e_g \quad (3-11-4)$$

The new current and voltage symbols of Eq. 3-11-4 denote increments in the variables. The constants r_p (incremental plate resistance) and g_m (incremental transconductance) are defined by the partial derivatives in Eq. 3-11-2 as evaluated at the operating point. Fig. 3-11-1b shows a circuit derived from Eq. 3-11-4. The circuit external to the tube has been replaced by its incremental model also—batteries are incremental short circuits, since they have a zero increment in voltage for an increment in current. The increments can be d-c changes or a-c signals. Once the circuit model has been constructed as in Fig. 3-11-1b, ordinary circuit analysis methods apply. The current generator, $g_m e_g$, is a dependent or controlled source and must be handled properly by relating e_g to the signal voltage, e_s . The solution of the model of Fig. 3-11-1b for the voltage gain $E_o(s)/E_s(s)$ proceeds as follows. A single node equation is written at the plate terminal.

$$E_o(s) \left[\frac{1}{r_p} + \frac{1}{R_L} \right] = -g_m E_g(s) = -g_m E_s(s) \quad (3-11-5)$$

$$\frac{E_o(s)}{E_s(s)} = \frac{-g_m}{\frac{1}{r_p} + \frac{1}{R_L}} = \frac{-g_m R_L}{1 + \frac{R_L}{r_p}} \quad (3-11-6)$$

Thus a voltage gain transfer function for the amplifier is given by Eq. 3-11-6. The amplifier can then be represented by the block of Fig. 3-11-1c for small signals that are slowly varying.

The basic incremental technique can be applied to many systems as is shown in later chapters. For example, the pentode tube amplifier model is the same as the triode if screen and suppressor voltages are held fixed. If the screen potential varies, the model is more complicated but can still be constructed.¹ At the higher frequencies the model of Fig. 3-11-1b must be complicated by adding interelectrode capacities between all pairs of electrodes; however, it still is a very useful model since most frequencies used in control systems are relatively low.

¹ E. J. Angelo, *Electronic Circuits*, McGraw-Hill Book Co., Inc., New York, 1958, Chapter 10.

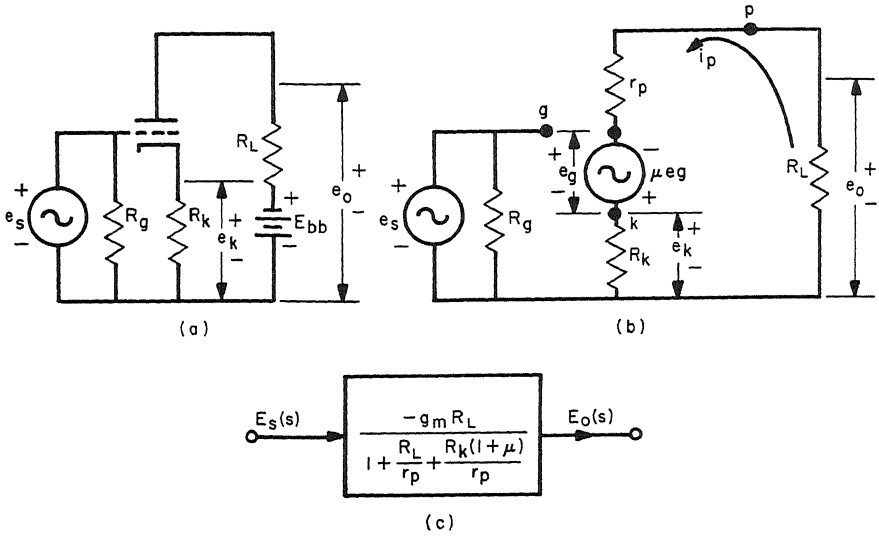


Fig. 3-11-2.

Example 3-11-1. The voltage gain transfer function for the amplifier of Fig. 3-11-2a is to be determined and a transfer function block diagram constructed. The circuit can be best analyzed by using a slightly modified equivalent circuit with a voltage source in series with r_p as obtained by multiplying Eq. 3-11-4 by r_p and using the new constant $\mu = r_p g_m$. Fig. 3-11-2b shows the resulting circuit model. The system equations are:

$$I_p(s)(R_L + r_p + R_k) = \mu E_g(s) \quad (3-11-7)$$

$$E_g(s) = E_s(s) - I_p(s)R_k \quad (3-11-8)$$

$$E_o(s) = -I_p(s)R_L \quad (3-11-9)$$

from which

$$\frac{E_o(s)}{E_s(s)} = \frac{-\mu R_L}{r_p + R_L + (1 + \mu)R_k} = \frac{-g_m R_L}{1 + R_L/r_p + (1 + \mu)R_k/r_p} \quad (3-11-10)$$

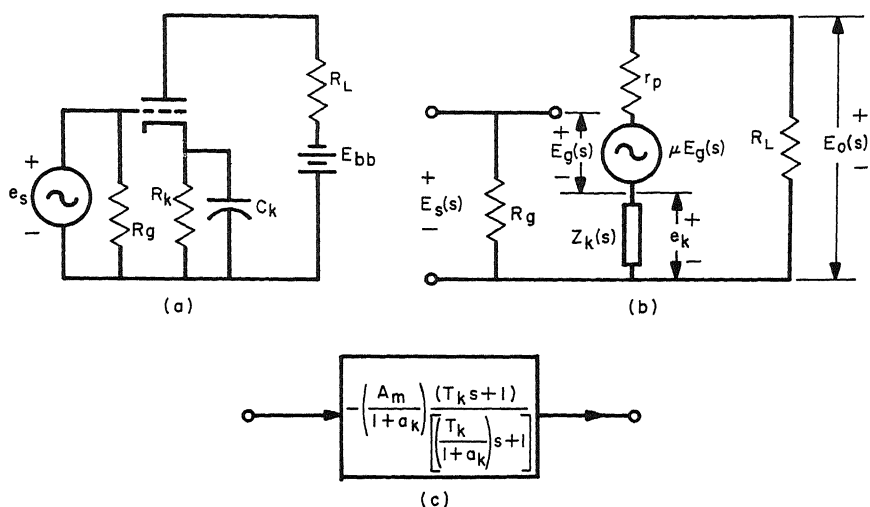
Fig. 3-11-2c shows the block diagram required.

Example 3-11-2. A block diagram is to be constructed for the voltage amplifier of Fig. 3-11-3a. The incremental circuit is shown in Fig. 3-11-3b. The simplest approach is to substitute

$$Z_k(s) = \frac{R_k(1/C_k s)}{R_k + 1/C_k s} = \frac{R_k}{R_k C_k s + 1} \quad (3-11-11)$$

in place of R_k in Eq. 3-11-10 and to rearrange factors as follows:

$$\frac{E_o(s)}{E_s(s)} = \left(\frac{-\mu R_L}{r_p + R_L} \right) \left(\frac{1}{1 + [(1 + \mu)R_k/(r_p + R_k)][1/(R_k C_k s + 1)]} \right) \quad (3-11-12)$$



Introducing

$$A_m = \frac{\mu R_L}{r_n + R_L} ; a_k = \frac{(1 + \mu) R_k}{r_n + R_k} ; T_k = C_k R_k \quad (3-11-13)$$

$$\frac{E_o(s)}{E_s(s)} = \left(\frac{-A_m}{1 + a_k} \right) \frac{T_k s + 1}{[T_k / (1 + a_k)] s + 1} \quad (3-11-14)$$

The extension of the incremental method transistor circuits is also easily accomplished;² however, limitations of space do not permit review of this subject here.

PROBLEMS

3-1. In Example 3-3-1, find the equation for e_c by starting with $I(s)$ of Eq. 3-6-6 in a similar manner to the way that e_L is determined in the example.

3-2. In Example 3-4-1, find the equation for i_c by starting with $E(s)$ of Eq. 3-4-5 in a similar manner to the way that i_L is determined in the example.

3-3. In the circuit in Fig. 3-P-3, S is closed until steady-state is reached. At $t = 0$, S is opened. Find the equation for the charge on the capacitor after $t = 0$, with the polarity reference as shown in the figure.

² *Ibid.*, Chapters 8 and 9.

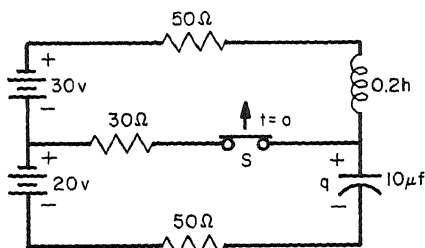


Fig. 3-P-3.

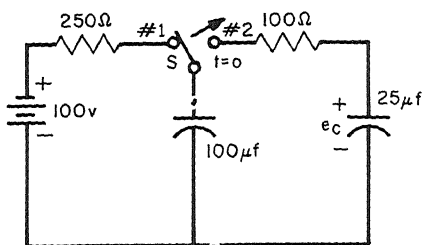


Fig. 3-P-4.

3-4. In the circuit in Fig. 3-P-4, S is closed in position 1 until steady state is reached. At $t = 0$, S is switched to position 2. Find the equation for the voltage across the $25 \mu f$ capacitor after $t = 0$, with the polarity reference shown.

3-5. In the circuit in Fig. 3-P-5, S is open until steady state is reached. At $t = 0$, S is closed. Find the equation for the current in the inductor after $t = 0$, in the direction shown.

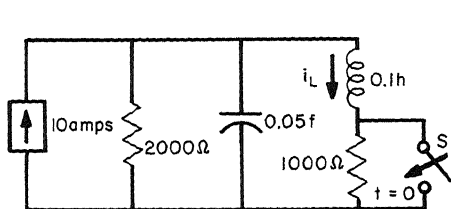


Fig. 3-P-5.

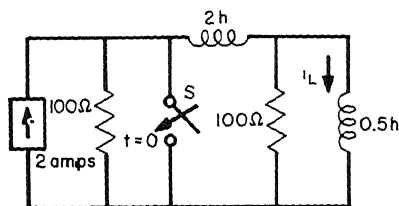


Fig. 3-P-6.

3-6. In the circuit in Fig. 3-P-6, S is open until steady state is reached. At $t = 0$, S is closed. Find the equation for the current in the 0.5 henry inductor after $t = 0$, in the direction shown.

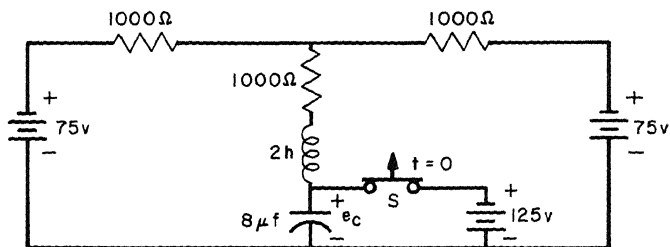


Fig. 3-P-7.

3-7. In the circuit in Fig. 3-P-7, S is closed until steady state is reached. At $t = 0$, S is opened. Find the equation for the voltage across the capacitor after $t = 0$, with the polarity reference shown.

3-8. Verify that $1/sC$ is the operational impedance of a capacitor.

3-9. Find the equation for e_p in Example 3-5-1 when the circuit is driven by $e = u_{-2}(t)$ in two different ways. First, substitute $E(s)$ into $E_p(s)$ and take the inverse Laplace transform. Second, integrate Eq. 3-5-35.

3-10. Find the equation for e_p in Example 3-5-1 when the circuit is driven by $e = u_0(t)$ in two different ways. First, substitute $E(s)$ into $E_p(s)$ and take the inverse transform. Second, differentiate Eq. 3-5-35.

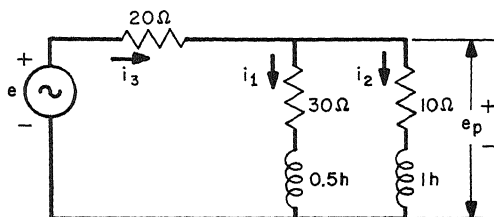


Fig. 3-P-11.

3-11. The circuit shown in Fig. 3-P-11 is at rest when the voltage e is applied to the circuit at $t = 0$.

(a) Operational methods are to be used to find the equations for $I_3(s)$, $I_2(s)$, $E_p(s)$ in terms of $I(s)$.

(b) The time functions i_3 , i_2 and e_p are to be found when the driving function is $e = u_0(t)$.

(c) The time functions i_3 , i_2 and e_p are to be found when the driving function is $e = u_{+1}(t)$ by substituting for $E(s)$ in the equations of Part (a) and obtaining the inverse transforms.

(d) Verify the solutions obtained in Part (c) by differentiating the solutions of Part (b).

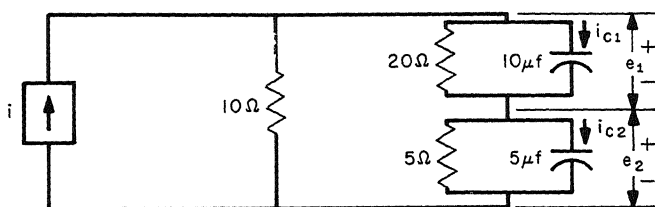


Fig. 3-P-12.

3-12. The circuit shown in Fig. 3-P-12 is at rest when the current i is applied at $t = 0$.

(a) Operational methods are to be used to find the equations for $E_1(s)$, $E_2(s)$ and $I_{C1}(s)$ in terms of $E(s)$.

(b) The time functions e_1 , e_2 and i_{c1} are to be found when the driving function is $i = u_{-1}(t)$.

(c) The time functions e_1 , e_2 and i_{c1} are to be found when the driving function is $i = u_{-2}(t)$ by substituting for $E(s)$ in the equations of Part (a) and obtaining the inverse transforms.

(d) Verify the solutions obtained in Part (c) by integrating the solutions of Part (b).

3-13. In the circuit of Fig. 3-P-11, the voltage e is given by $e = \cos 20t$. The voltage is applied to the circuit and is left until steady state is reached. Find i_{3ss} , i_{2ss} and e_{pss} by using the $s = j\omega$ technique. Check these results by using conventional a-c circuit theory.

3-14. In the circuit of Fig. 3-P-12, the current i is given by $i = \cos 5,000t$. The current is applied to the circuit and is left until steady state is reached. Find e_{1ss} , e_{2ss} and i_{C1ss} by using the $s = j\omega$ technique. Check these results by using conventional a-c circuit theory.

3-15. In the circuit of Fig. 3-P-11, the driving function is $e = [\cos 20t]u_{-1}(t)$ and the complete solutions for i_3 , i_2 and e_p are to be found. Compare the steady-state components of the solutions with those found in Prob. 3-13.

3-16. In the circuit of Fig. 3-P-12, the driving function is $i = [\cos 5,000t]u_{-1}(t)$ and the complete solutions for e_1 , e_2 and i_{C1} are to be found. Compare the steady-state components of the solutions with those found in Prob. 3-14.

3-17. In example 3-7-1 the driving function $e = [\cos 10t]u_{-1}(t)$ is applied to the circuit of Fig. 3-6-1a and the response function for i_2 is found as Eq. 3-7-8. The driving function e and the response are to be differentiated. The new response for i_2 is to be verified by differentiating e and by solving for i_2 after applying the new driving function to the circuit. Note that the original e has a discontinuity at $t = 0$.

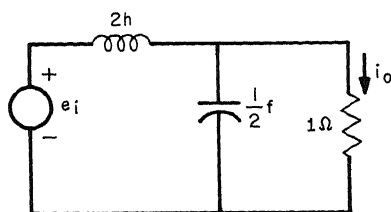


Fig. 3-P-18.

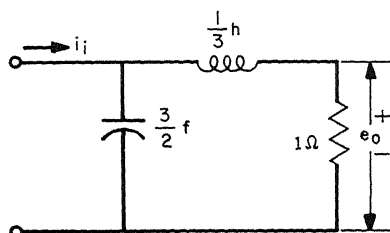


Fig. 3-P-19.

3-18. Find the transfer function and draw the block diagram for the circuit of Fig. 3-P-18.

3-19. Find the transfer function and draw the block diagram for the circuit of Fig. 3-P-19.

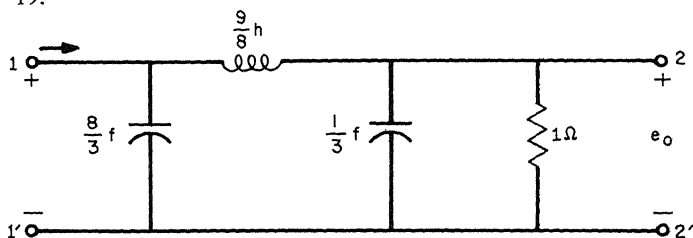


Fig. 3-P-20.

3-20. Find the transfer function and draw the block diagram for the circuit of Fig. 3-P-20.

3-21. Repeat Prob. 3-20 except that current i_i is supplied at terminals 2-2' and voltage e_o is obtained at terminals 1-1'. Compare the answers to Probs. 3-20 and 3-21. What theorem does this comparison verify?

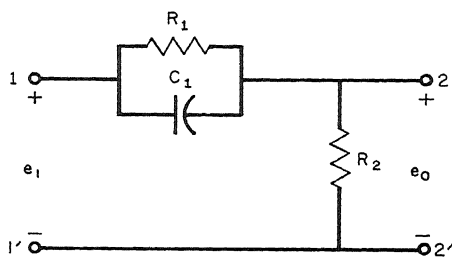


Fig. 3-P-22.

3-22. Find the transfer function and draw the block diagram for the circuit of Fig. 3-P-22.

3-23. Repeat Prob. 3-22 except that voltage e_i is applied at terminals 2-2' and voltage e_o is obtained at terminals 1-1'. Why does this answer differ from that of Prob. 3-22?

3-24. Find the transfer function and draw the block diagram for the circuit of Fig. 3-P-24.

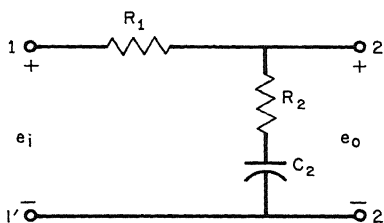


Fig. 3-P-24.

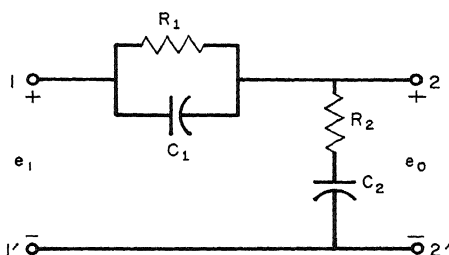


Fig. 3-P-25.

3-25. Find the transfer function and draw the block diagram for the circuit of Fig. 3-P-25.

3-26. Two networks of the type shown in Fig. 3-P-22 are connected in tandem with an isolation amplifier having unity gain connecting the output of the first network to the input of the second. Find the transfer function for the combination and draw the block diagram.

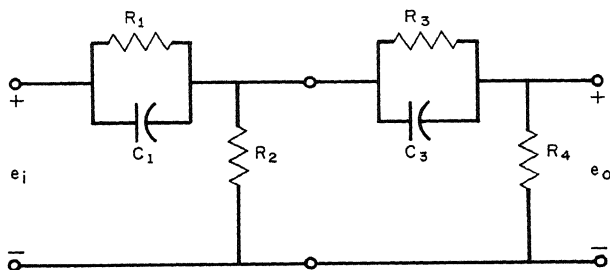


Fig. 3-P-27.

3-27. Find the transfer function and draw the block diagram for the circuit of Fig. 3-P-27. Compare the results with those found in Prob. 3-26.

3-28. Repeat Prob. 3-26 using the network shown in Fig. 3-P-24.

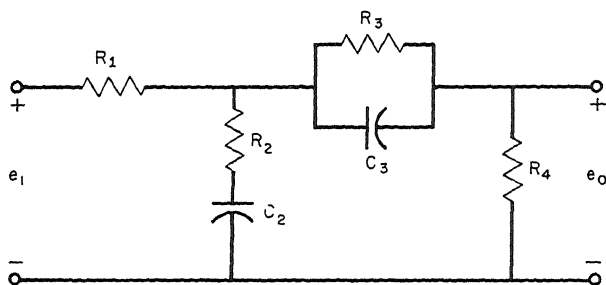


Fig. 3-P-29.

3-29. Find the transfer function and draw the block diagram for the circuit of Fig. 3-P-29.

3-30. The circuit of Fig. 3-P-22 is loaded with a resistor R_3 . Find the transfer function for the combination.

3-31. The circuit of Fig. 3-P-24 is loaded with a series R_3 - C_3 circuit. The product R_2C_2 is equal to the product R_3C_3 . Find the transfer function for the combination.

3-32. Transform the initial conditions of Prob. 3-3 into voltage sources and solve by using operational methods.

3-33. Transform the initial conditions of Prob. 3-4 into voltage sources and solve by using operational methods.

3-34. Transform the initial conditions of Prob. 3-5 into current sources and solve by using operational methods.

3-35. Transform the initial conditions of Prob. 3-6 into current sources and solve by using operational methods.

3-36. Transform the initial conditions of Prob. 3-7 into voltage sources and solve by using operational methods.

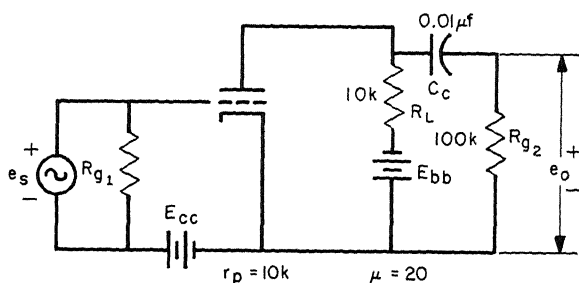


Fig. 3-P-37.

3-37. Find the incremental voltage gain transfer function for the amplifier in Fig. 3-P-37 at low and medium frequencies.

3-38. Find the incremental voltage gain for the amplifier in Fig. 3-P-38 at low and medium frequencies.

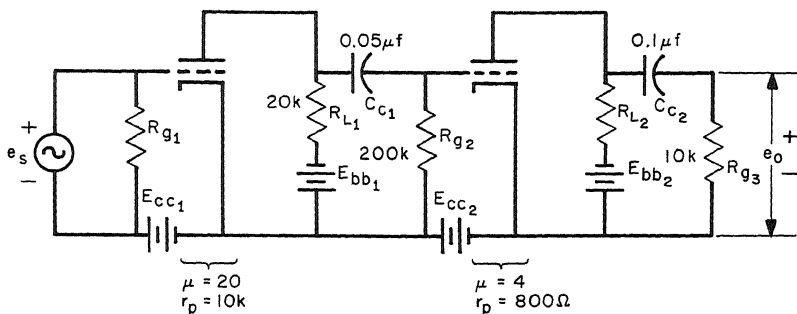


Fig. 3-P-38.

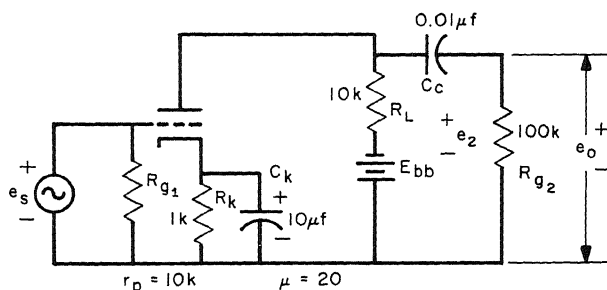


Fig. 3-P-39.

3-39. Find the incremental voltage gain transfer function for the amplifier in Fig. 3-P-39 at low and medium frequencies. Consider this case two ways.

(a) Use the complete circuit.

(b) Find $E_2(s)$ by neglecting the loading of C_c and R_{g2} and then find $E_o(s)$ by voltage dividing action of C_c and R_{g2} . Compare results with Prob. 3-37.

MECHANICAL AND ELECTROMECHANICAL SYSTEMS

4-1. Mechanical Parameters and Variables. The solution of many mechanical system problems can be obtained by using linear lumped mechanical constants just as electrical circuits in Chapter 3 are solved by using R , L , and C . The basic mechanical parameters can be divided into two classes, one involving translation and one involving rotation. The parameters for translational motion are rigid mass, M , viscous friction, f_v , and the linear spring constant, K_s . For rotational motion the mass is replaced by mass moment of inertia, J , the viscous friction coefficient becomes D , and the spring constant remains K_s although it changes dimensions. Each of these constants is examined in detail in later sections.

The variables generally encountered in mechanical systems are force (torque), position, velocity, acceleration, and time. The following three articles review the relationships and procedures necessary for solving the problems encountered in many mechanical systems.

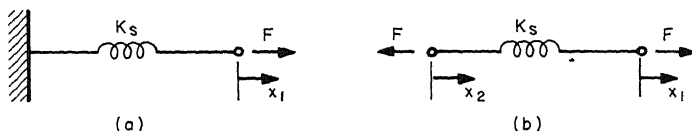


Fig. 4-2-1.

4-2. The Linear Spring. Fig. 4-2-1 shows the symbol that is used for the linear spring when one terminal is fixed to rigid structure. Without application of an external force the spring will assume a so-called free or unstretched length. The position variable x_1 is generally chosen so that it will be zero when the spring is unstretched or when the system is at rest. The direction of positive increasing value for x_1 is arbitrarily assigned as shown in Fig. 4-2-1a. Since only one end of the spring can move in the case being considered, only one position variable is needed. A positive direction for force is also shown in Fig. 4-2-1a. The equation relating the external force, F , the position x_1 , and the spring constant, K_s , can now be written for a linear spring as

$$F = K_s x_1 \quad (4-2-1)$$

The deflection of the spring from its rest position is directly proportional to the external force applied. Compression of the spring would require a negative value of x_1 for the case shown in Fig. 4-2-1a and this in turn yields a numerically negative force, F , in Eq. 4-2-1.

If both ends of the spring are allowed to move, the net elongation or compression of the spring depends upon two position variables. A new second position variable must be introduced to define the position of the end of the spring that was formerly fixed. Thus x_2 is introduced in Fig. 4-2-1b with both x_1 and x_2 generally being taken as zero when the spring is unstretched. The external force, F , now depends upon the actual elongation, $(x_1 - x_2)$, of the spring. The force-displacement relationship for this case is

$$F = K_s(x_1 - x_2) \quad (4-2-2)$$

If x_1 is greater than x_2 , the force required is positive; whereas if x_1 is less than x_2 , the applied force must be negative and the spring is actually compressed.

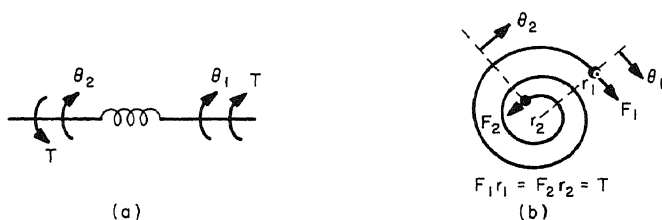


Fig. 4-2-2.

For rotational systems the spring is commonly represented by either of the symbols shown in Fig. 4-2-2. Two angular position variables, θ_1 and θ_2 , are chosen so that they are both zero when the spring is free or at rest and the directions of positive increasing θ_1 and θ_2 are arbitrarily assigned. A positive direction of torque is also assumed as shown in Fig. 4-2-2. The equation

$$T = K_s(\theta_1 - \theta_2) \quad (4-2-3)$$

can now be written since the net angular deflection of the spring is directly proportional to the external torque. The rotational spring constant need not represent an actual spiral spring but may represent a shaft which has an effective spring constant and must be treated as a spring.

Several additional factors related to springs should be considered before proceeding to a second mechanical parameter. The spring constant for translational motion has dimensions such as pounds/foot or ounces/inch while for rotational motion the dimensions are foot-pounds/radian and inch-ounces/degree. The spring stores potential energy and dissipates

no average power. From basic energy and power relations the following equations

$$\text{Work} = \int_0^x F dx = \int_0^{x_1} K_s(x_1 - x_2) d(x_1 - x_2) \quad (4-2-4)$$

$$= \frac{K_s}{2} (x_1 - x_2)^2 = \text{Stored energy} \quad (4-2-5)$$

$$\text{Power} = \frac{d(\text{Work})}{dt} = K_s(x_1 - x_2) \frac{d(x_1 - x_2)}{dt} \quad (4-2-6)$$

can be written. Any force or torque applied at one end is instantly transmitted directly to the other end.

4-3. Viscous Friction or Damping. The type of friction used in linear differential equation studies is called viscous friction. While other types of friction exist in mechanical systems, they very frequently result in forces much smaller than those due to viscous friction or to other damping effects. In addition, the force expressions required to account for other types of friction are nonlinear and thereby provide an additional basis for neglecting all but viscous friction in a first linear analysis of a system.

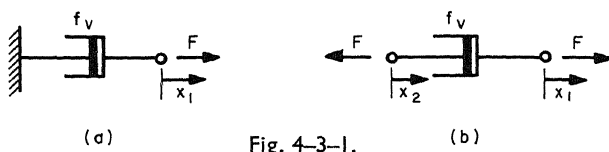


Fig. 4-3-1.

Fig. 4-3-1a shows the symbol that is used for the viscous friction element when one terminal is fixed. A displacement variable, x_1 , is chosen in addition to a positive force variable F as shown in Fig. 4-3-1a. The equation for force can now be written as

$$F = f_v \frac{dx_1}{dt} \quad (4-3-1)$$

Eq. 4-3-1 actually defines the viscous friction coefficient, f_v , and in fact, describes the nature of this type of friction. The force transmitted by the viscous element is directly proportional to the relative velocity between ends of the element. The constant f_v is the proportionality constant.

If both ends of the viscous friction element are free to move, a second displacement variable, x_2 , is needed. The choice of direction for x_2 is arbitrary. This is always the case when choosing position variables. In the situation being considered, Fig. 4-3-1b illustrates a typical choice for x_1 and x_2 . The relation between force and velocity now becomes

$$F = f_v \frac{d(x_1 - x_2)}{dt} \quad (4-3-2)$$

This indicates that, if the velocity of the x_1 end of the element is greater than the velocity of the x_2 end, a positive force is required. No force is transmitted by the element when the two ends have the same velocity.

The same symbol is used to represent viscous friction in both rotational and translational motion; however, in rotational motion the viscous friction coefficient is designated by D and it now relates torque to relative angular velocity. Fig. 4-3-2 illustrates the viscous friction element and

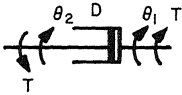


Fig. 4-3-2.

$$T = D \frac{d(\theta_1 - \theta_2)}{dt} \quad (4-3-3)$$

a typical choice of position variables, θ_1 , θ_2 , and torque, T , for rotational motion. Dimensionally D differs from f_v , since the dimensions of f_v are pounds/(foot/second), ounces/(inch/second), etc. while D has dimensions such as foot-pounds/(radian/second) and ounce-inches/(degree/second).

As mentioned previously, two other types of friction are also encountered in physical systems and these lead to difficulties in solving the system differential equations that include their effects. The first of these types of friction is called coulomb friction and can be accounted for by a constant magnitude force always resisting motion; the second type is called "stiction" and is the frictional force that must be overcome in setting a body in motion.¹

In contrast to the spring, viscous friction does not result in storage of energy, but instead in dissipation of energy in the form of heat. The power and energy equations needed for the viscous friction element can be found by the same methods employed for finding the spring energy and power equations.

$$\text{Work} = \int_0^x F dx = \int_0^{x_1 - x_2} f_v \frac{d(x_1 - x_2)}{dt} d(x_1 - x_2) \quad (4-3-4)$$

$$\text{Power} = \frac{d(\text{Work})}{dt} = f_v \left[\frac{d(x_1 - x_2)}{dt} \right]^2 \quad (4-3-5)$$

For rotational motion, D , θ , and T replace f_v , x , and F respectively in Eq. 4-3-4 and Eq. 4-3-5.

4-4. Mass and Moment of Inertia. The last parameter to be studied for translatory motion is that of mass, M , as introduced in basic physics. Newton's second law states the relationship needed for writing a force equation. Thus,

$$F = \frac{d(\text{Momentum})}{dt} = \frac{d(mv)}{dt} \quad (4-4-1)$$

¹ Robert A. Bruns and Robert M. Saunders, *Analysis of Feedback Control Systems*, McGraw-Hill Book Co., Inc., New York, 1955, pp. 23-25.

This equation can be rewritten in the following form when the mass involved is constant.

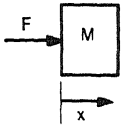


Fig. 4-4-1.

$$\text{Force} = (\text{Mass})(\text{Acceleration}) = ma \quad (4-4-2)$$

Hence

$$F = M \frac{d^2x}{dt^2} \quad (4-4-3)$$

Fig. 4-4-1 shows both the symbol Eq. 4-4-3 used for mass and the choice for the one position variable needed to locate the mass when motion is possible in one direction only.

The force, F , may be the sum of a large number of component forces. In determining signs for force components, the following rule suffices: A force which tends to produce greater positive displacement is taken as a positive force in Eq. 4-4-3.

For rotational systems the mass moment of inertia, J , replaces the mass parameter of translational systems. This parameter relates applied torque to angular acceleration as follows:

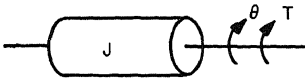


Fig. 4-4-2.

$$T = J \frac{d^2\theta}{dt^2} \quad (4-4-4)$$

Fig. 4-4-2 shows both the symbol used for the moment of inertia and an assumed direction for angular position, θ , and torque, T . The rule above concerning component forces can be extended to cover component torques. That is, if a component torque tends to cause a positive angular motion, the torque is positive.

A review of the dimensions of M and J would be useful at this point. In the English system of units, mass has dimensions such as pounds/(feet, second²) or slugs while in the metric system the dimensions are grams and kilograms. Care must be exercised in distinguishing between a gram of mass and a gram of weight. In the English system of unit, the mass moment of inertia has dimensions such as foot-pounds/(radians/second²) or slug-feet². In the metric system, the mass moment of inertia has dimensions such as gram-centimeters/(radians/second²).

A consideration of the power and energy relations for mass and mass moment of inertia reveals that these elements store kinetic energy by virtue of their velocities and potential energy by virtue of their position above a reference plane. No energy is dissipated as heat. The power and energy equations for mass and mass moment of inertia are

$$\text{KE} = \int_0^x M \frac{d^2x}{dt^2} dx = \int_0^t M \frac{d^2x}{dt^2} \frac{dx}{dt} dt = \frac{M}{2} \left(\frac{dx}{dt} \right)^2 \quad (4-4-5)$$

or

$$KE = \frac{J}{2} \left(\frac{d\theta}{dt} \right)^2 \quad (4-4-6)$$

and

$$\text{Power} = M \frac{d^2 x}{dt^2} \frac{dx}{dt} \quad (4-4-7)$$

or

$$\text{Power} = J \frac{d^2 \theta}{dt^2} \frac{d\theta}{dt} \quad (4-4-8)$$

4-5. Mechanical System Equations. Many mechanical systems can be analyzed by application of the force and torque relations of Arts. 4-2, 4-3, and 4-4. Each mass or moment of inertia in the system is isolated in a free body diagram showing all forces or torques. Newton's second law is then applied to each free body. In the process, positive displacement variables must be assumed for the ends of each spring, each viscous friction element, and each free body.

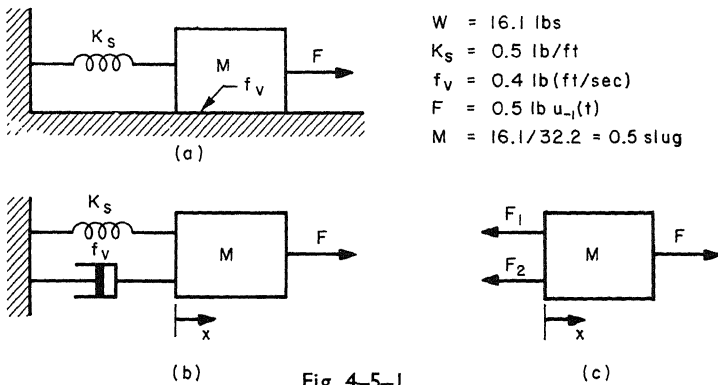


Fig. 4-5-1.

Example 4-5-1. As an example of this procedure the simple spring, mass, friction system of Fig. 4-5-1a is analyzed to obtain the transfer function relating the position of the mass to the applied force. It is assumed that the mass moves in only one direction; therefore, only one position variable is needed in this problem. The variable x is chosen so that it is zero when the spring is unstretched and the system is at rest. Fig. 4-5-1b shows the system in schematic form; however, it must be observed that in this diagram force due to the weight of the body has been omitted since it does not affect motion in the x direction. Fig. 4-5-1c shows a free body diagram of the mass, again neglecting weight. Force F is an external applied force while force F_1 represents the effect of the spring and F_2 represents the effect of the viscous friction. Newton's second law, as applied to the free body of Fig. 4-5-1c, gives the relation

$$M \frac{d^2 x}{dt^2} = F - F_1 - F_2 \quad (4-5-1)$$

An expression for F_1 is found by considering the actual spring force as the mass moves in the positive x direction. This procedure gives the relation

$$F_1 = K_s x \quad (4-5-2)$$

The equation for F_2 is found by considering the actual frictional force as the mass moves with a positive velocity.

$$F_2 = f_v \frac{dx}{dt} \quad (4-5-3)$$

Eq. 4-5-1 can now be rewritten to include Eq. 4-5-2 and Eq. 4-5-3.

$$M \frac{d^2 x}{dt^2} = F - K_s x - f_v \frac{dx}{dt} \quad (4-5-4)$$

With the system initially at rest the application of a step function of applied force cannot produce an initial velocity, $dx(0)/dt$, or an initial position, $x(0)$. The applied force initially accelerates the mass. Initial conditions can therefore be neglected when taking the Laplace transform of Eq. 4-5-4. In any event, initial conditions cannot be included in an ordinary transfer function. The Laplace transform of Eq. 4-5-4 therefore appears as follows:

$$Ms^2 X(s) = F(s) - K_s X(s) - f_v s X(s) \quad (4-5-5)$$

Eq. 4-5-5 can be rearranged in a manner permitting formation of the desired transfer function.

$$(Ms^2 + f_v s + K_s) X(s) = F(s) \quad (4-5-6)$$

From Eq. 4-5-6 the transfer function relating the output position to the applied force is

$$G(s) = \frac{X(s)}{F(s)} = \frac{1}{Ms^2 + f_v s + K_s} \quad (4-5-7)$$

With respect to input force and output position, the system of Fig. 4-5-1a can now be represented by the block diagram of Fig. 4-5-2. The transfer function concept is then used to express the output position as

$$X(s) = G(s)F(s) \quad (4-5-8)$$

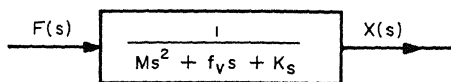


Fig. 4-5-2.

For the particular numerical values given in Fig. 4-5-1a, Eq. 4-5-8 can be written

$$X(s) = \frac{1}{0.5s^2 + 0.4s + 0.5} \cdot \frac{0.5}{s} \quad (4-5-9)$$

$$= \frac{1}{s(s^2 + 0.8s + 1)} \quad (4-5-10)$$

To find the time variation of x it is necessary to find the inverse transform of Eq. 4-5-10. This process is simplified if $X(s)$ is written as

$$X(s) = \frac{1}{s[(s + 0.4)^2 + (0.92)^2]} \quad (4-5-11)$$

The desired time function can now be found by using the table of Laplace transform pairs in the Appendix.

$$x(t) = [1 + 1.09e^{-0.4t} \sin(0.92t - 113.6^\circ)] \text{ ft} \quad (4-5-12)$$

When the velocity rather than the position of the mass is of interest, the velocity transfer function can be found by multiplying Eq. 4-5-7 by s . This is the equivalent of a differentiation for rest conditions. This procedure results in the following equation for $V(s)$, the Laplace transform of velocity.

$$V(s) = sG(s)F(s) \quad (4-5-13)$$

$$= \frac{1}{(s + 0.4)^2 + (0.92)^2} \quad (4-5-14)$$

The time equation for velocity as obtained from the inverse transform is

$$v(t) = 1.09e^{-0.4t} \sin 0.92t \quad (4-5-15)$$

The same answer can be obtained by differentiating Eq. 4-5-12 and by simplifying the result.

Example 4-5-2. The system used in Example 4-5-1 comprises a second example when it is initially in motion and when there is no external driving force. For this case the initial velocity is taken as -1 ft/second and the initial position is assumed to be 0.5 ft. Since normal transfer functions do not involve initial conditions, a straightforward use of the Laplace transform of Eq. 4-5-4 is required. In this example the applied force, F , is zero. The direct Laplace transform of Eq. 4-5-4 with the above conditions is

$$Ms^2 X(s) - Msx(0) - M \frac{dx(0)}{dt} = -K_s X(s) - f_v s X(s) + f_v x(0) \quad (4-5-16)$$

Since $dx(0)/dt$ and $x(0)$ are known quantities, they are grouped together to give

$$(Ms^2 + f_v s + K_s)X(s) = M \frac{dx(0)}{dt} + (Ms + f_v)x(0) \quad (4-5-17)$$

The initial condition terms now act as driving functions with the initial position being equivalent to an impulse of force, $f_v x(0)$, and a doublet of force, $M s x(0)$. The initial velocity is also equivalent to an impulse of force, $M dx(0)/dt$. In other words, exactly the same equation results if the system is initially at rest and if impulse and doublet driving forces, as described above, are applied. A rearrangement of Eq. 4-5-17 and the substitution of numerical values as given previously allows $X(s)$ to be expressed as

$$X(s) = \frac{(0.5)(-1) + (0.5s + 0.4)(0.5)}{0.5s^2 + 0.4s + 0.5} \quad (4-5-18)$$

$$= \frac{(0.5)(s - 1.2)}{s^2 + 0.8s + 1} = \frac{(0.5)(s - 1.2)}{(s + 0.4)^2 + (0.92)^2} \quad (4-5-19)$$

By using the transform tables again the time varying position, $x(t)$ is found to be

$$x(t) = 1.0e^{-0.4t} \sin(0.92t + 150^\circ) \text{ ft} \quad (4-5-20)$$

A check of the information contained in the transfer function of Eq. 4-5-7 shows that its poles are the same as those in Eq. 4-5-19. Since the poles of the transfer function, $G(s)$, determine the response modes such as damping rates and damped angular velocities, the transfer function contains the basic information concerning system response. No great amount of system information is actually lost by neglecting initial conditions and by studying only the transfer functions of control system components. This fact becomes even more apparent when it is realized that the responses due to the various initial conditions can be obtained by considering the system initially at rest and by then exciting it with various impulse functions. Therefore most of this text utilizes the transfer function approach to system analysis.

Example 4-5-3. The system used in Example 4-5-1 is studied when it is excited or driven by a steady-state sinusoidal force, $F = 0.5 \sin t$ in order to find the steady-state variation of the position of the mass.

In Chapter 3 the phasor transform is introduced as a special form of the Laplace transform. This specialized transform is valid only for a study of the steady-state performance of a system with sinusoidal excitation. It provides no information regarding the transient part of the response. As pointed out in Chapter 3 the phasor or steady-state transfer function is formed by letting $s = j\omega$ in the original Laplace transform transfer function. This limits s to imaginary values whereas in general s can be a complex number, $s = \sigma + j\omega$. It follows that such a restriction results in a transfer function of more limited use.

By substituting $s = j\omega = j1$ in Eq. 4-5-7 the phasor or steady-state transfer function for this example is obtained.

$$G(j\omega) = \frac{X(j\omega)}{F(j\omega)} = \frac{1}{(K_s - M\omega^2) + jf_r\omega} \quad (4-5-21)$$

$$= \frac{1}{(0.5 - 0.5) + j0.4} \quad (4-5-22)$$

$$= \frac{1}{j0.4} = 2.5 \angle -90^\circ \quad (4-5-23)$$

Eq. 4-5-23 gives the ratio of the sinusoidal amplitude of $x(t)$ in the steady state to the amplitude of $F(t)$ and also gives the relative phase angle between the time response and the exciting force. Therefore $x_{ss}(t)$ can be found from

$$X(j\omega) = G(j\omega)F(j\omega) = (2.5 \angle -90^\circ)(0.5 \angle 0^\circ) \quad (4-5-24)$$

$$x_{ss}(t) = 1.25 \sin(t - 90^\circ) = -1.25 \cos t \quad (4-5-25)$$

If the transient part of the response is desired, it can be obtained by using the direct Laplace transform of the force $F(t)$ as follows:

$$x(s) = \frac{1}{(0.5s^2 + 0.4s + 0.5)} \frac{(0.5)}{(s^2 + 1)} \quad (4-5-26)$$

$$= \frac{1}{(s^2 + 1)[(s + 0.4)^2 + (0.92)^2]} \quad (4-5-27)$$

$$= \frac{-1.25s}{s^2 + 1} + \frac{1.25s + 1}{[(s + 0.4)^2 + (0.92)^2]} \quad (4-5-28)$$

Since the first term in Eq. 4-5-28 gives the sinusoidal steady-state response, the second term must give the transient part of the response.

$$x(t)_{\text{tran}} = 1.36e^{-0.4t} \sin(0.92t + 66.5^\circ) \quad (4-5-29)$$

The evaluation of the transient response is not usually necessary in problems involving sinusoidal excitation. Therefore, only the steps through Eq. 4-5-25 are normally required.

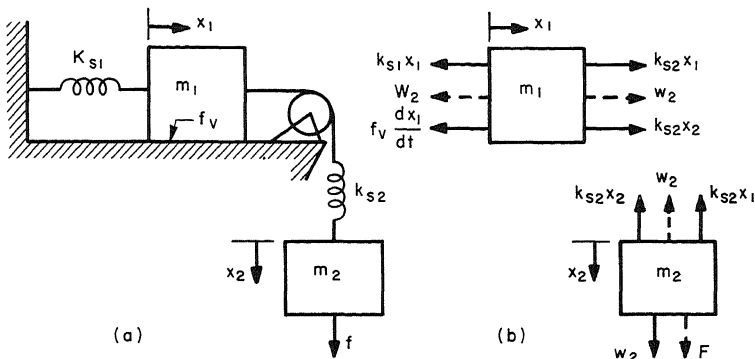


Fig. 4-5-3.

Example 4-5-4. A more complicated problem involving two masses introduces the idea of superposition of motions. Fig. 4-5-3a shows the system to be studied. The transfer function relating the position of mass M_1 to applied force, F , is to be found. Masses M_1 and M_2 are assumed to have motion in only one direction. Each mass therefore requires one position variable to define its motion. Variables x_1 and x_2 are arbitrarily chosen as shown in Fig. 4-5-3a; however they are chosen to be zero when the system is at rest but not when the springs are unstretched. This means that initially the springs are stretched enough to balance the force due to the weight of body M_2 thereby establishing static equilibrium. The dotted forces labeled W_2 in Fig. 4-5-3b are included to show the actual existence of W_2 in spite of the fact that they do not need to appear in the system equations when x_1 and x_2 are chosen as discussed above.

The isolation of M_1 and M_2 as free bodies is the next step in obtaining the desired transfer function, $X_1(s)/F(s)$. The various component forces acting on M_1 and M_2 must be determined. A scheme of superimposing motions is useful here. The effects of allowing x_1 to increase are first determined while x_2 is held fixed at zero. This leads immediately to forces $K_{s1}x_1$ and $f_v dx_1/dt$ which act in the directions indicated in the M_1 free body diagram of Fig. 4-5-3b. Spring K_{s2} will be relaxed as x_1 increases while x_2 is fixed so that the full force W_2 is no longer transmitted by

spring K_{s2} . Effectively, then, a force $K_{s2}x_1$ must be included in the free body diagrams for M_1 and M_2 as shown in Fig. 4-5-3b. This force should really be grouped with W_2 in both free bodies to obtain the real force transmitted by K_{s2} when x_1 moves and x_2 is zero. All of the effects of allowing x_1 to increase by itself have now been determined and included in the appropriate free body diagrams.

The next step necessitates a study of the effects of allowing x_2 to increase while x_1 is held at zero. The only force component that results in this case is $K_{s2}x_2$ which is due to stretching spring K_{s2} . This force acts on M_1 and M_2 in the directions indicated in Fig. 4-5-3b. Since the cable over the pulley cannot transmit compression, the accuracy of the free body diagrams is limited to cases in which

$$K_{s2}(x_1 - x_2) < W_2 \quad (4-5-30)$$

Newton's second law can now be applied by using the free body diagrams developed above to give

$$M_1 \frac{d^2 x_1}{dt^2} = [W_2 - K_{s2}(x_1 - x_2)] - W_2 - K_{s1}x_1 - f_v \frac{dx_1}{dt} \quad (4-5-31)$$

$$M_2 \frac{d^2 x_2}{dt^2} = F + W_2 - [W_2 - K_{s2}(x_1 - x_2)] \quad (4-5-32)$$

The force components that tend to cause motion in the positive x_1 and x_2 directions are taken as positive forces.

In order to obtain the desired transfer function, $X_1(s)/F(s)$ the Laplace transform of Eq. 4-5-31 and Eq. 4-5-32 must be formed assuming rest conditions in the system. A rearrangement of these equations is also necessary in order to solve them simultaneously for $X_1(s)$ in terms of $F(s)$. The final relations are

$$[M_1 s^2 + f_v s + (K_{s1} + K_{s2})]X_1(s) - K_{s2}X_2(s) = 0 \quad (4-5-33)$$

$$-K_{s2}X_1(s) + [M_2 s^2 + K_{s2}]X_2(s) = F(s) \quad (4-5-34)$$

Eqs. 4-5-33 and 4-5-34 can be solved by determinants or some other suitable scheme to give

$$\frac{X_1(s)}{F(s)} = \frac{K_{s2}}{[M_1 s^2 + f_v s + (K_{s1} + K_{s2})][M_2 s^2 + K_{s2}] - K_{s2}^2} \quad (4-5-35)$$

An equivalent form for Eq. 4-5-35 can be written by expanding the denominator as follows:

$$\frac{X_1(s)}{F(s)} = \frac{K_{s2}}{M_1 M_2 s^4 + M_2 f_v s^3 + [M_2(K_{s1} + K_{s2}) + M_1 K_{s2}]s^2 + K_{s2} f_v s + K_{s1} K_{s2}} \quad (4-5-36)$$

The restrictions on the validity of the transfer function of Eq. 4-5-36 must be checked after $x_1(t)$ and $x_2(t)$ are found for any given input. Several of these restrictions have been merely implied. Mass M_1 must not hit the pulley, stretch spring K_{s1} beyond its elastic limit, or compress K_{s1} beyond its linear range. Mass M_1 and M_2 together must not move so as to violate the inequality of Eq. 4-5-30 or to stretch spring K_{s2} beyond its elastic limit. Within these limits Eq. 4-5-36 is correct.

By solving Eqs. 4-5-33 and 4-5-34 for $X_2(s)$ as a function of $F(s)$, the transfer function $X_2(s)/F(s)$ can be obtained. The denominator of this relation is the same as the denominator of Eq. 4-5-36. This indicates that the same basic response modes (damping rates, natural frequencies, etc.) are present in $x_2(t)$ as in $x_1(t)$; only the amplitudes and relative phase angles are different.

Example 4-5-5. The rotational mechanical system of Fig. 4-5-4a is analyzed to obtain the time varying position of the cylindrical body J_1 when a sudden torque, T , is applied to the system which is initially at rest. It is assumed that the shaft

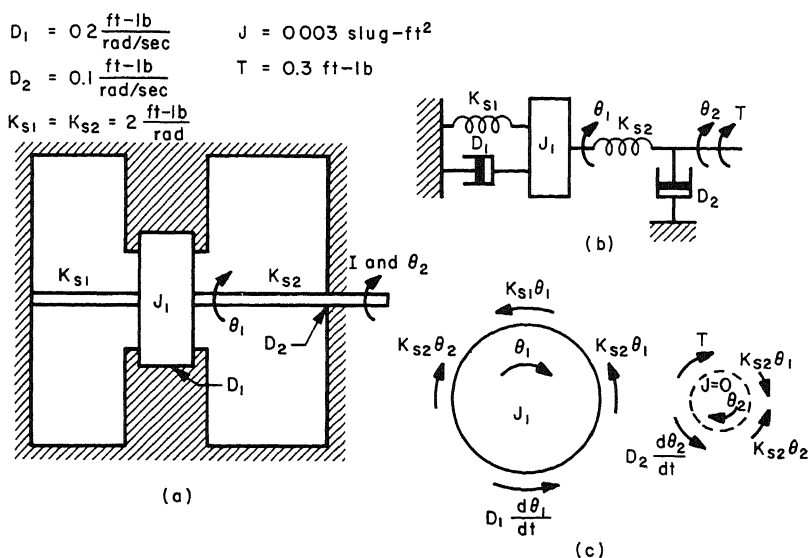


Fig. 4-5-4.

deflects enough to require representation by springs K_{s1} and K_{s2} and that viscous friction exists at the surfaces indicated by the arrows for D_1 and D_2 . Fig. 4-5-4b shows a schematic of the system and indicates a choice for the two position variables, θ_1 and θ_2 , that are required in the solution of this problem. While no moment of inertia is assumed to exist at the right-hand end of the shaft in Fig. 4-5-4a, a position variable is nevertheless required to determine the motion of this point.

The isolation of free bodies is the next step in determining the system equations. To simplify the determination of torque components, a free body with zero moment of inertia is drawn at the location of position variable θ_2 . The superposition of motions as used in the previous example is also useful in this problem. Body J_1 is assumed to be moving in the positive θ_1 direction while θ_2 is held at zero. This results in the torque components $K_{s1}\theta_1$, $K_{s2}\theta_1$, and $D_1 \frac{d\theta_1}{dt}$ acting on body J_1 as indicated in Fig. 4-5-4c. Spring K_{s2} also couples the torque $K_{s2}\theta_1$ to the other free body in Fig. 4-5-4c. With θ_1 returned to zero θ_2 is now allowed to increase. Torque components $K_{s2}\theta_2$ and $D_2 \frac{d\theta_2}{dt}$ oppose this motion while the component $K_{s2}\theta_2$ is coupled to body J_1 as shown in Fig. 4-5-4c. The system equations can now be written by using the free body diagrams.

$$J_1 \frac{d^2 \theta_1}{dt^2} = -D_1 \frac{d\theta_1}{dt} - K_{s2}(\theta_1 - \theta_2) - K_{s1}\theta_1 \quad (4-5-37)$$

$$0 = T - K_{s2}(\theta_2 - \theta_1) - D_2 \frac{d\theta_2}{dt} \quad (4-5-38)$$

A rearrangement to obtain symmetry of form results in

$$J_1 \frac{d^2 \theta_1}{dt^2} + D_1 \frac{d\theta_1}{dt} + (K_{s2} + K_{s1})\theta_1 - K_{s2}\theta_2 = 0 \quad (4-5-39)$$

$$-K_{s2}\theta_1 + D_2 \frac{d\theta_2}{dt} + K_{s2}\theta_2 = T \quad (4-5-40)$$

The Laplace transforms of Eqs. 4-5-39 and 4-5-40 are now formed, considering the fact that the system is initially at rest:

$$[J_1 s^2 + D_1 s + (K_{s1} + K_{s2})]\Theta_1(s) - K_{s2}\Theta_2(s) = 0 \quad (4-5-41)$$

$$-K_{s2}\Theta_1(s) + [D_2 s + K_{s2}]\Theta_2(s) = I/s \quad (4-5-42)$$

In order to find the desired position variation Eqs. 4-5-41 and 4-5-42 must be solved simultaneously for $\Theta_1(s)$. The simplified result is

$$\Theta_1(s) = \frac{K_{s2} T}{s[J_1 D_2 s^3 + (D_1 D_2 + J_1 K_{s2})s^2 + (D_1 K_{s2} + D_2 K_{s1} + D_2 K_{s2})s + K_{s1} K_{s2}]} \quad (4-5-43)$$

The substitution of numerical values from Fig. 4-5-4a in Eq. 4-5-43 is necessary in order to complete the solution for $\theta_1(t)$. While numerical values could have been used throughout the problem, the use of letter parameters allows an easy dimensional check on the coefficient of each power of s . For example, the s^2 coefficient in Eq. 4-5-43 has the dimensions

$$D_1 D_2 \left(\frac{\text{ft} \cdot \text{lbs}}{\text{radian} \cdot \text{sec}} \right)^2 + J_1 K_{s2} \left(\frac{\text{ft} \cdot \text{lbs}}{\text{radian} \cdot \text{sec}^2} \right) \left(\frac{\text{ft} \cdot \text{lbs}}{\text{radian}} \right) \quad (4-5-44)$$

or $(\text{ft}^2 \cdot \text{lbs}^2 \cdot \text{sec}^2)/\text{radian}^2$. When the dimensions of s , which are seconds^{-1} , are included, each term in the denominator polynomial should be consistent.

$\Theta_1(s)$ can be expressed as follows:

$$\Theta_1(s) = (2)(0.3)/s \{ (3 \times 10^{-3})(0.2)s^3 + [(0.2)(0.1) + (3 \times 10^{-3})(2)]s^2 + [(0.2)(2) + (0.1)(2) + (0.2)(2)]s + (2)(2) \} \quad (4-5-45)$$

$$= \frac{1000}{s(s^3 + 43.3s^2 + 1333s + 6667)} \quad (4-5-46)$$

$$= \frac{1000}{s(s + 6)[(s + 18.6)^2 + (27.6)^2]} \quad (4-5-47)$$

The inverse Laplace transform of Eq. 4-5-46, as found by using the table in the Appendix, is

$$\theta_1(t) = [0.15 - 0.18e^{-6t} + 0.036e^{-18.6t} \sin(27.6t + 122^\circ)] \text{ radian} \quad (4-5-48)$$

4-6. Gear Trains. Like the mechanical elements considered in earlier sections, the gear train appears in many control system problems. Gear trains are used for speed reduction, for matching driving members to given loads for most efficient power transfer or for maximum load acceleration, and for many other functions.²

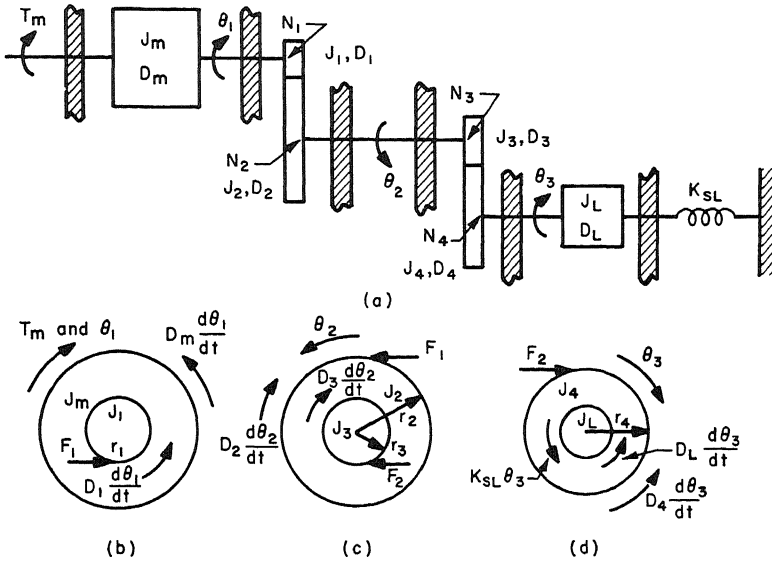


Fig. 4-6-1.

A typical gear train is pictured in Fig. 4-6-1a. Each gear and pinion is given a moment of inertia and a viscous friction drag. A driving torque and load are also included. Several of the parameters might be negligible or completely absent in practical cases, especially the load spring constant. The free bodies of Fig. 4-6-1b, c and d can be obtained by a process of isolating gears while making use of material in the preceding sections on mechanical parameters. Newton's second law is now applied to each of the free bodies to give

$$(J_1 + J_m) \frac{d^2 \theta_1}{dt^2} = T_m(t) - F_1 r_1 - (D_m + D_1) \frac{d\theta_1}{dt} \quad (4-6-1)$$

$$(J_2 + J_3) \frac{d^2 \theta_2}{dt^2} = F_1 r_2 - F_2 r_3 - (D_2 + D_3) \frac{d\theta_2}{dt} \quad (4-6-2)$$

$$(J_L + J_4) \frac{d^2 \theta_3}{dt^2} = F_2 r_4 - (D_4 + D_L) \frac{d\theta_3}{dt} - K_{SL} \theta_3 \quad (4-6-3)$$

² John G. Truxal, *Control Engineers Handbook*, McGraw-Hill Book Co., Inc., New York, 1958, Section 13-1.

The following auxiliary equations are written utilizing the geometry of the system:

$$\theta_1 r_1 = \theta_2 r_2 \quad (4-6-4)$$

$$\theta_2 r_3 = \theta_3 r_4 \quad (4-6-5)$$

A rearrangement of Eqs. 4-6-4 and 4-6-5 allows the introduction of the gear ratio, n , as well as the gear angular velocities, ω_1 , ω_2 , and ω_3 :

$$\frac{\theta_1}{\theta_2} = \frac{\omega_1}{\omega_2} = \frac{r_2}{r_1} = \frac{N_2}{N_1} = n_1 \quad (4-6-6)$$

$$\frac{\theta_2}{\theta_3} = \frac{\omega_2}{\omega_3} = \frac{r_4}{r_3} = \frac{N_4}{N_3} = n_2 \quad (4-6-7)$$

$$\frac{\theta_1}{\theta_3} = \frac{\omega_1}{\omega_3} = n_1 n_2 = n \quad (4-6-8)$$

The gear ratio as defined here and as used in this text is the ratio of the diameter or number of teeth of the driven gear to the diameter or number of teeth of the driving gear. For speed-reducing gear trains n is greater than one. Eqs. 4-6-1, 4-6-2, and 4-6-3 may be combined by eliminating F_1 and F_2 and by introducing various gear ratios. If θ_1 is retained, the result of this reduction process is

$$\begin{aligned} & [(J_1 + J_m) + \frac{1}{n_1^2}(J_2 + J_3) + \frac{1}{n^2}(J_4 + J_L)] \frac{d^2 \theta_1}{dt^2} \\ & + [(D_1 + D_m) + \frac{1}{n_1^2}(D_2 + D_3) + \frac{1}{n^2}(D_4 + D_L)] \frac{d \theta_1}{dt} \\ & + \frac{K_{s1}}{n^2} \theta_1 = T_m(t) \end{aligned} \quad (4-6-9)$$

The output or load position, θ_3 , could have been retained instead of θ_1 .

$$\begin{aligned} & [n^2(J_1 + J_m) + n_2^2(J_2 + J_3) + (J_4 + J_L)] \frac{d^2 \theta_3}{dt^2} \\ & + [n^2(D_1 + D_m) + n_2^2(D_2 + D_3) + (D_4 + D_L)] \frac{d \theta_3}{dt} \\ & + K_{s1} \theta_3 = n T_m(t) \end{aligned} \quad (4-6-10)$$

The important feature of these last two equations is the way in which the effects of the mechanical parameters transfer through the gear train. In moving the effect of a parameter from one shaft to another shaft nearer the output, the parameter is multiplied by the square of the gear ratio involved. The applied torque, which is a system variable, is multiplied by the gear ratio. Parameters that are transferred toward the input shaft are divided by the square of the gear ratio as shown in Eq. 4-6-9. When

moments of inertia, friction, and spring constants exist at more than one shaft in a system containing a gear train, effective values of these parameters must be considered when system equations are written.

Example 4-6-1. A common problem encountered in control systems having gear trains is that of applying a torque to an input shaft and finding the output shaft position. A transfer function relating $\Theta_3(s)$ to $T_m(s)$ is needed. To simplify the equations that follow, J_{eff_1} and D_{eff_1} are defined as

$$J_{\text{eff}_1} = n^2(J_1 + J_m) + n_2^2(J_2 + J_3) + (J_4 + J_L) \quad (4-6-11)$$

$$D_{\text{eff}_1} = n^2(D_1 + D_m) + n_2^2(D_2 + D_3) + (D_4 + D_L) \quad (4-6-12)$$

and

$$J_{\text{eff}_1} = \frac{J_{\text{eff}_1}}{n^2} \quad (4-6-13)$$

$$D_{\text{eff}_1} = \frac{D_{\text{eff}_1}}{n^2} \quad (4-6-14)$$

$$K_{s_{\text{eff}_1}} = \frac{K_{s_i}}{n^2} \quad (4-6-15)$$

With the system of Fig. 4-6-1a at rest the Laplace transform of Eq. 4-6-10 is

$$(J_{\text{eff}_1} s^2 + D_{\text{eff}_1} s + K_{s_i})\Theta_3(s) = nT_m(s) \quad (4-6-16)$$

The desired transfer function is

$$G(s) = \frac{\Theta_3(s)}{T_m(s)} = \frac{n}{J_{\text{eff}_1} s^2 + D_{\text{eff}_1} s + K_{s_i}} \quad (4-6-17)$$

The system of Fig. 4-6-1a can be represented by the block diagram of Fig. 4-6-2.

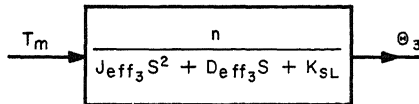


Fig. 4-6-2.

Example 4-6-2. The problem of Example 4-6-1 is modified to introduce the input shaft position and to illustrate an alternate way of accounting for the over-all gear train. The first step involves finding a transfer function between input shaft position, θ_1 , and input torque, T_m . This transfer function is provided by the Laplace transform of Eq. 4-6-9,

$$\frac{\Theta_1(s)}{T_m(s)} = \frac{1}{J_{\text{eff}_1} s^2 + D_{\text{eff}_1} s + K_{s_{\text{eff}_1}}} \quad (4-6-18)$$

The relationship in Eq. 4-6-8 provides the remaining equation needed to represent the gear train, as shown in Fig. 4-6-3. This over-all block diagram is equivalent to the single block of Fig. 4-6-2.

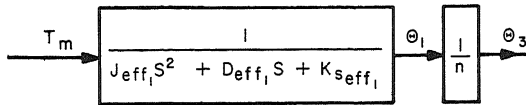


Fig. 4-6-3.

4-7. Mechanical Differential. The use of feedback requires a device for comparing an input signal with a feedback signal. In most control systems this device must perform the operation of subtraction and must provide the resulting difference as an output signal. When the signals involved are mechanical positions or velocities, the mechanical differential is commonly used. Fig. 4-7-1a and b shows the symbols

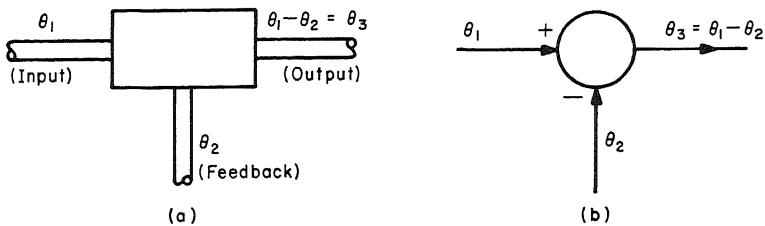


Fig. 4-7-1.

that are used in this text to represent an idealized version of the mechanical differential. Fig. 4-7-1a is used in mechanical schematics and Fig. 4-7-1b is used in one line schematics. This idealized version has no appreciable inertia or friction and it is assumed that an appreciable moment of inertia does not exist on the main input shaft nor on the output shaft. This eliminates the problem of transferring moment of inertia effects through mechanical differentials. The equation that relates the position variables can be written as

$$\theta_3 = \theta_1 - \theta_2 \quad (4-7-1)$$

In this equation θ_1 is normally the main input, θ_2 is a secondary input or feedback signal, and θ_3 is the output or difference signal. Later chapters consider other summing devices that behave in a similar manner.

4-8. D-c Generators. This article introduces the first electro-mechanical device that is analyzed to obtain a specific transfer function. The d-c generator whose transfer function is to be found is used to provide controlled power to actuate various system components. In this and in the following articles, some knowledge of the theory and construction of generators and motors is assumed. However, most of the equations require only introductory magnetic field concepts.

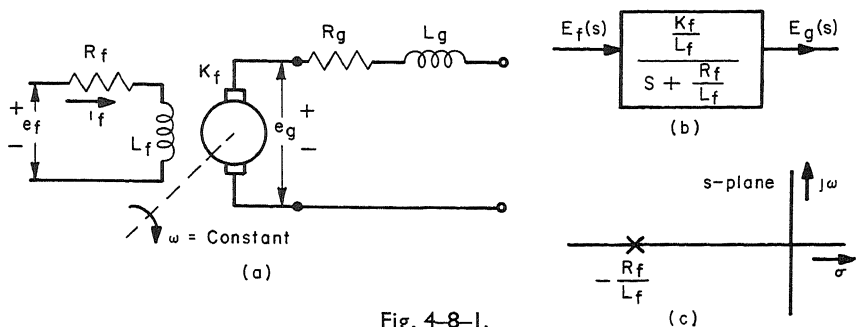


Fig. 4-8-1.

A convenient starting point in the analysis is the schematic representation of a d-c generator as shown in Fig. 4-8-1a. The field of the generator is represented by the circuit which contains R_f and L_f in Fig. 4-8-1a. Its current establishes the main magnetic field in the machine. The circuit representing the armature contains R_g and L_g . The armature is driven at a constant speed by an external prime mover. The generated voltage results when the armature conductors cut the magnetic field flux. For a linear magnetic circuit the voltage induced in the armature is directly proportional to the field current. The resulting drop is expressed by the equation

$$e_g = K_f i_f \quad (4-8-1)$$

in which K_f is a constant, e_g is the open-circuit armature voltage, and i_f is the field current. From circuit theory the field equation is

$$i_f R_f + L_f \frac{di_f}{dt} = e_f \quad (4-8-2)$$

If the relationship between the applied field voltage, e_f , and the open-circuit armature voltage, e_g , is desired, Eqs. 4-8-1 and 4-8-2 provide adequate information. The Laplace transforms of these two equations as needed for a transfer function are

$$E_g(s) = K_f I_f(s) \quad (4-8-3)$$

$$(R_f + L_f s) I_f(s) = E_f(s) \quad (4-8-4)$$

The desired transfer function can be written by combining Eqs. 4-8-3 and 4-8-4.

$$\frac{E_g(s)}{E_f(s)} = \frac{K_f/L_f}{(s + R_f/L_f)} \quad (4-8-5)$$

Fig. 4-8-1b is a block diagram that contains the information in Eq. 4-8-5 and that can therefore represent the generator. Eq. 4-8-5 indicates that the relationship between open-circuit output voltage and field voltage is characterized by a single pole in the Laplace or s -plane. This pole is

always on the negative real axis and therefore gives a decaying exponential transient term in the time response to an excitation. The significant dynamic characteristics of the basic unloaded d-c generator can be represented by a plot of the s -plane showing the single pole at $s = -R_f/L_f$. Fig. 4-8-1c is such a pole-zero plot. This plot does not contain the numerator constant of the transfer function of Eq. 4-8-5.

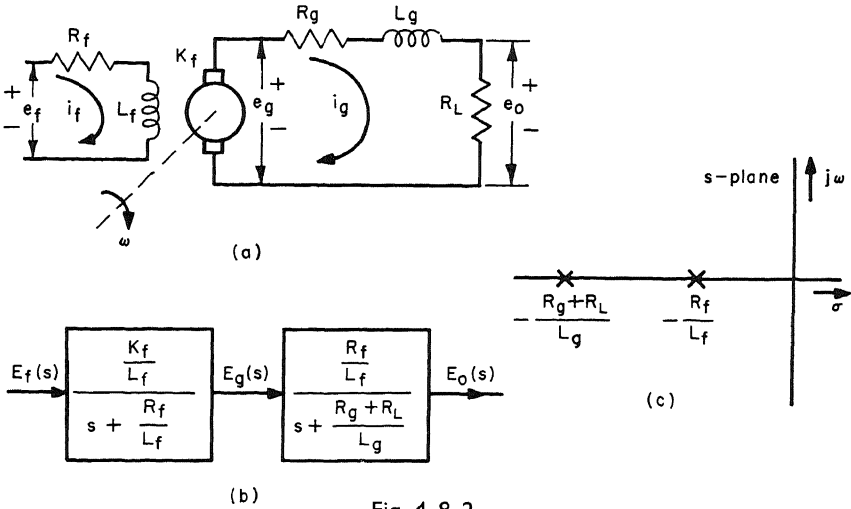


Fig. 4-8-2.

Example 4-8-1. When the generator armature circuit is loaded, as shown in Fig. 4-8-2a, the armature resistance and inductance and the load circuit must be considered in the analysis. A transfer function is to be found relating the output voltage, e_o , of Fig. 4-8-2a to the field voltage. Since e_g is already related to e_f , this step need not be repeated. The two output circuit equations

$$(R_g + R_L + L_g s)I_g(s) = E_g(s) \quad (4-8-6)$$

$$E_o(s) = R_L I_g(s) \quad (4-8-7)$$

can be obtained directly by using the operational impedances of Chapter 3. The transfer function relating $E_o(s)$ to $E_g(s)$ results from a combination of Eqs. 4-8-6 and 4-8-7.

$$\frac{E_o(s)}{E_g(s)} = \frac{R_L/L_g}{s + (R_g + R_L)/L_g} \quad (4-8-8)$$

The product of $E_g(s)/E_f(s)$ from Eq. 4-8-5 and $E_o(s)/E_g(s)$ from Eq. 4-8-8 is the desired over-all transfer function, $E_o(s)/E_f(s)$. Fig. 4-8-2b shows how the original block diagram of Fig. 4-8-1b must be modified to apply to the loaded generator in this problem. A single block which contains the product of the two individual transfer functions could be used. The s -plane representation in Fig. 4-8-1c must also be changed to fit the case under consideration. A second pole must be added at $s = -(R_g + R_L)/L_g$ as shown in Fig. 4-8-2c. Two decaying exponential transient terms appear in the time response of the loaded generator of Fig. 4-8-2a.

4-9. The Amplidyne. A specialized d-c generator called the amplidyne has been developed to provide very large power gain and good dynamic characteristics. This machine is really two generator stages on one frame. A detailed study of the magnetic flux and the generated voltage in a generator must be carried out in order to develop the theory of the amplidyne. Fig. 4-9-1a shows a reasonably detailed schematic diagram

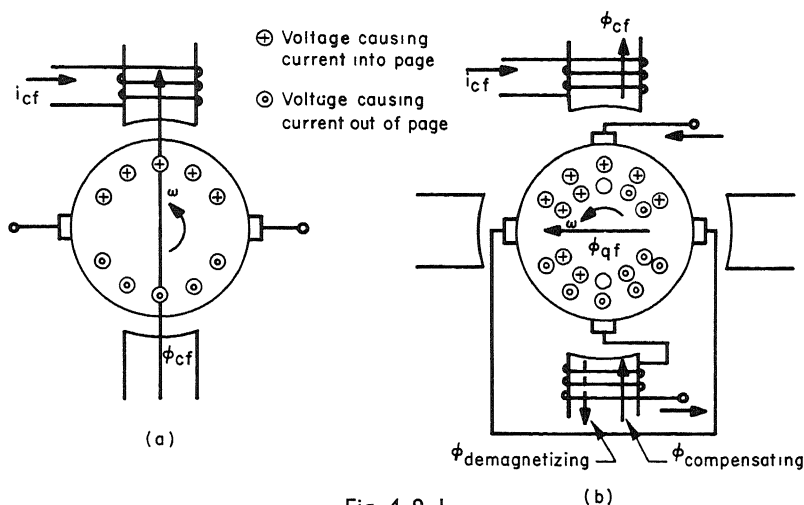


Fig. 4-9-1.

of a basic d-c generator. A typical field winding, armature conductors, field flux ϕ_{cf} , direction of armature rotation, and direction of induced armature voltage are all illustrated. The basic field in Fig. 4-9-1a is called the control field in the amplidyne. The normal armature output terminals are shorted together as shown in Fig. 4-9-1b. This causes a large current to flow in the armature, thereby establishing a second magnetic flux ϕ_{qf} which is perpendicular to the control field flux as shown in Fig. 4-9-1b. This flux is called the quadrature field flux. The outer ring of circles indicates the direction for the current that establishes ϕ_{qf} . The armature conductors cut ϕ_{qf} as well as ϕ_{cf} and a second component voltage is induced in the armature winding by the quadrature field flux. The direction of this voltage is shown by the inner ring of circles. A special set of brushes has been added in Fig. 4-9-1b to bring out the voltage generated by the quadrature field. In order to prevent current flow in this new output circuit from establishing a flux ϕ_{demag} that would counteract or demagnetize the original control field, a compensating field is added. This field, as shown in Fig. 4-9-1b, is connected in the output circuit and develops a compensating flux, ϕ_{comp} , that cancels ϕ_{demag} . This cancellation is possible since both of these flux components are proportional to load current. The control field of this machine serves as the input controlling a large power output.

A suitable transfer function for the amplidyne can be obtained by using a procedure similar to that used for the ordinary generator. If linear magnetic circuits are assumed and if mutual inductance between fields is negligible, the following operational equations can be written for the control field circuit and for the quadrature field open-circuit voltage, $E_{qf}(s)$.

$$E_{cf}(s) = (R_{cf} + L_{cf}s) I_{cf}(s) \quad (4-9-1)$$

$$E_{qf}(s) = K_{qc} I_{cf}(s) \quad (4-9-2)$$

The constant K_{qc} is determined by the magnitude of the quadrature field induced voltage per ampere of control field current. The other quantities, E_{cf} , R_{cf} , L_{cf} and I_{cf} are the control field applied voltage, resistance, inductance, and current respectively. In a similar manner two equations for the quadrature field can be written.

$$E_{qf}(s) = (R_{qf} + L_{qf}s) I_{qf}(s) \quad (4-9-3)$$

$$E_g(s) = K_{gq} I_{qf}(s) \quad (4-9-4)$$

The generated voltage constant relating open-circuit output voltage, $E_g(s)$, to quadrature field current, $I_{qf}(s)$, is K_{gq} . The parameters R_{qf} and L_{qf} are the quadrature field resistance and inductance. Eqs. 4-9-1 through 4-9-4 when combined reduce to the following transfer function for an amplidyne,

$$\frac{E_g(s)}{E_{cf}(s)} = \frac{K_{gq} K_{qc} / L_{cf} L_{qf}}{(s + R_{cf} / L_{cf})(s + R_{qf} / L_{qf})} \quad (4-9-5)$$

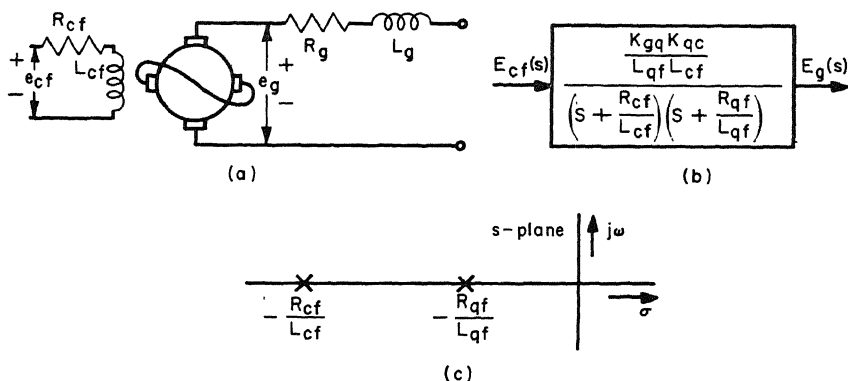


Fig. 4-9-2.

Figs. 4-9-2a, b, and c present a summary of the amplidyne by showing its schematic, its block diagram, and its s -plane pole-zero characteristic.

4-10. Synchro Units. A flexible and accurate means of transmitting and manipulating mechanical shaft position information is needed in many control systems. A number of devices called synchros are extensively employed in the above application. Basically, synchros are analogous to systems of a-c electromagnets. Fig. 4-10-1a and b show two pivoted

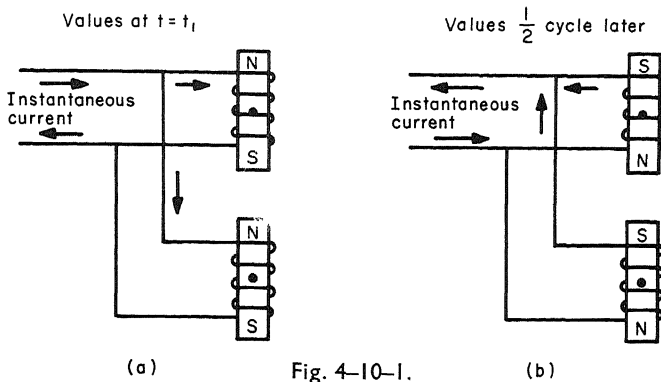


Fig. 4-10-1.

a-c electromagnets that illustrate their tendency to align themselves and to hold this alignment in spite of the time variation of magnetic flux.

There are five basic synchro units and these are generally called synchro motors, synchro generators, synchro control transformers, synchro differential generators, and synchro differential motors.

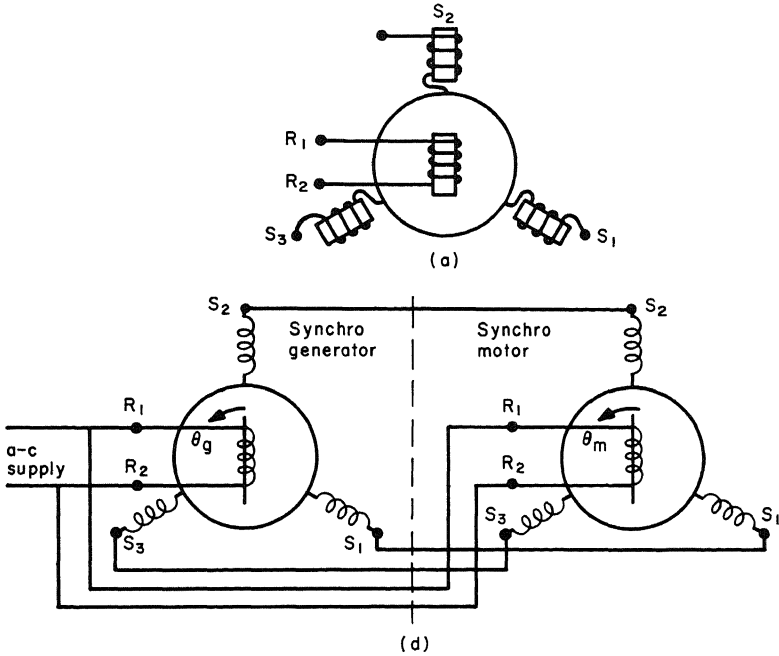


Fig. 4-10-2.

The synchro motor can be pictured as shown in Fig. 4-10-2a. The three wye-connected windings are on the stator of the machine while the single center coil is wound on the rotor. When various combinations of a-c voltages are applied to the three stator windings, a resultant magnetic flux is established. By varying the stator voltage magnitudes the resultant flux can be rotated to any angular position. If voltage is also supplied to the rotor, it will tend to align itself with the resultant stator field.

The three stator voltages for the synchro motor are supplied by another member of the synchro family, the synchro generator. Fig. 4-10-2b shows a schematic circuit for a synchro generator connected to a synchro motor. The circuit similarity between the two units is readily apparent. The synchro motor, however, has considerably higher mechanical damping to prevent excessive spinning. When in operation the rotor of the synchro motor tends to follow the position of the generator rotor. This action is due to torques that exist when the various resultant magnetic fields in the synchro motor and generator are not aligned. The alignment torque in the motor is given by

$$T = K \sin (\theta_g - \theta_m) \quad (4-10-1)$$

where θ_g and θ_m are the angles of the generator and motor shaft with respect to a common point on the stator of the two machines. The axis of stator coil S_2 on each unit is a commonly used reference point. A similar torque tends to pull the synchro generator rotor into alignment with the motor rotor; however, θ_g is generally held at a given input angle and is not moved by this torque. As a linear approximation for small values of $\theta_g - \theta_m$, Eq. 4-10-1 can be written as

$$T \approx K(\theta_g - \theta_m) \quad (4-10-2)$$

Another combination of synchro units that is of basic importance in control systems is shown in Fig. 4-10-3. The synchro control transformer is similar to a synchro motor. However, the control transformer rotor is

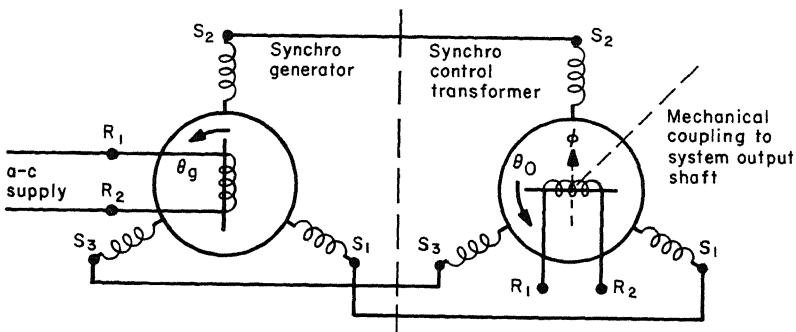


Fig. 4-10-3.

not connected to the a-c supply. Instead the control transformer rotor is mechanically coupled to the system output shaft as shown in Fig. 4-10-3. An output voltage exists at terminals R_1 and R_2 of the control transformer whenever the axis of its rotor is not perpendicular to the resultant magnetic flux established by currents in the stator windings. The resultant stator flux rotates with the rotor of the generator. Therefore, the control transformer output voltage is an error voltage depending upon the difference in the synchro generator rotor position and the system output shaft position which is coupled to the rotor of the control transformer. The error voltage amplitude is

$$e_{os} = K_1 \sin(\theta_g - \theta_o) \quad (4-10-3)$$

in which θ_g is the generator rotor position and θ_o is the position of the control transformer as shown in Fig. 4-10-3.

θ_o has its zero reference shifted 90° from the zero reference for θ_g . The actual error voltage is an a-c voltage of supply frequency with the amplitude

$$e_o(t) = K_1 \sin[\theta_g(t) - \theta_o(t)] \sin 2\pi f_{a-c} t \quad (4-10-4)$$

determined by Eq. 4-10-3. This is a type of amplitude or envelope modulation. The real signal information is contained in the envelope which must be recovered for use in control systems having d-c motors, generators, and other d-c components. For linear analysis, Eq. 4-10-3 is approximated as

$$e_{os}(t) \approx K_1[\theta_g(t) - \theta_o(t)] \quad (4-10-5)$$

when $\theta_g - \theta_o$ is small. The Laplace transform of Eq. 4-10-5

$$E_{os}(s) \approx K_1[\Theta_g(s) - \Theta_o(s)] \quad (4-10-6)$$

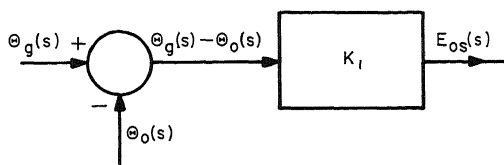


Fig. 4-10-4.

can be used to construct the block diagram of Fig. 4-10-4 as a suitable approximate representation of the synchro generator-control transformer combination. It must be remembered that Eq. 4-10-6 and Fig. 4-10-4 are valid only for small angular errors and that only the envelope part of the error voltage is represented.

The interested reader is referred to other sources for a fuller introduction to the various members of the synchro family and a more detailed study of the synchro generator, motor, and control transformer.^{3, 4}

³ U.S. Navy Synchros, Bureau of Ordnance publication OP1303 (15 April 1958).

⁴ John E. Gibson and Franz B. Tuteur, *Control System Components*, McGraw-Hill Book Co., Inc., New York, N.Y., 1958, Chapter 5.

4-11. D-c Servomotors (Armature Controlled).⁵ Most control systems involve components that require combinations of electrical and mechanical equations. The d-c servomotor is such a component. The analysis of this device proceeds from basic relations for electrical conductors moving in magnetic fields, together with electrical circuit and mechanical equations. Space does not permit a review of all the basic physics involved. Therefore, a starting point that assumes some knowl-

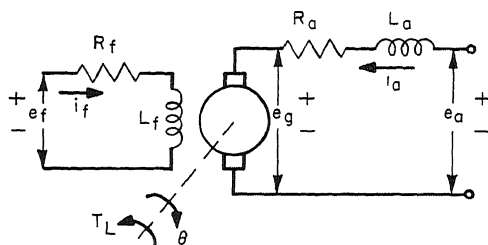


Fig. 4-11-1.

edge of electrical machines is used. Fig. 4-11-1 shows the schematic for a d-c motor. The various symbols used in connection with this figure are

e_a = armature terminal voltage

i_a = armature current

R_a = armature resistance

L_a = armature inductance

R_f = field resistance

L_f = effective field inductance

e_f = motor field voltage

i_f = motor field current

e_g = emf developed in armature winding due to cutting magnetic flux

ϕ = magnetic flux crossing air gap between field poles and armature

T_D = electrically developed torque

T_L = load torque

K_1, K_2, K_3 = constants

θ = armature position variable

J_e = effective load moment of inertia at motor shaft (includes J_{motor})

D_e = effective load viscous friction at motor shaft

K_{se} = effective load spring constant at motor shaft

⁵ G. J. Thaler and W. A. Stein, "Transfer Function and Parameter Evaluation for DC Servomotors," *Trans. AIEE*, 75 (January, 1956): 410-417.

The following equations can be written for the field of a linear d-c motor:

$$R_f i_f + L_f \frac{di_f}{dt} = e_f \quad (4-11-1)$$

$$K_1 i_f = \varphi \quad (4-11-2)$$

$$K_2 \varphi \frac{d\theta}{dt} = e_g \quad (4-11-3)$$

The armature circuit equation is

$$i_a R_a + L_a \frac{di_a}{dt} + e_g = e_a \quad (4-11-4)$$

while the electrically developed driving torque can be written as

$$K_3 \varphi i_a = T_D \quad (4-11-5)$$

A general mechanical load made up of an effective moment of inertia, J_e , an effective viscous friction, D_e , and an effective spring constant, K_{se} is assumed. The load torque equation is

$$K_{se} \theta + D_e \frac{d\theta}{dt} + J_e \frac{d^2 \theta}{dt^2} = T_L \quad (4-11-6)$$

The speed and position of an armature controlled d-c motor are determined by varying the armature voltage, e_a . The field current, i_f , is maintained constant thereby causing the flux φ in Eq. 4-11-2 to be constant. Eqs. 4-11-3 and 4-11-5 can be modified for this case as follows,

$$e_g = K_{av} \frac{d\theta}{dt} = K_{av} \omega \quad (4-11-7)$$

$$T_D = K_T i_a \quad (4-11-8)$$

A transfer function relating the input armature voltage to the armature position is needed for use in the block diagram of the armature controlled d-c motor. As a first step in obtaining this transfer function Eqs. 4-11-7 and 4-11-8 are used to rewrite Eq. 4-11-4 as follows

$$e_a = \frac{R_a}{K_T} T_D + \frac{L_a}{K_T} \frac{dT_D}{dt} + K_{av} \frac{d\theta}{dt} \quad (4-11-9)$$

The load torque, T_L , in Eq. 4-11-6 must equal the developed torque, T_D , thus allowing T_D to be eliminated in Eq. 4-11-9.

$$e_a = \frac{R_a}{K_T} \left(J_e \frac{d^2 \theta}{dt^2} + D_e \frac{d\theta}{dt} + K_{se} \theta \right) + \frac{L_a}{K_T} \frac{d}{dt} \left(J_e \frac{d^2 \theta}{dt^2} + D_e \frac{d\theta}{dt} + K_{se} \theta \right) + K_{av} \frac{d\theta}{dt} \quad (4-11-10)$$

$$\begin{aligned}
 = & \frac{L_a J_e}{K_T} \frac{d^3 \theta}{dt^3} + \left(\frac{R_a J_e}{K_T} + \frac{L_a D_e}{K_T} \right) \frac{d^2 \theta}{dt^2} + \left(\frac{R_a D_e}{K_T} + \frac{L_a K_{se}}{K_T} + K_{av} \right) \frac{d\theta}{dt} \\
 & + \frac{R_a K_{se}}{K_T} \theta \quad (4-11-11)
 \end{aligned}$$

Eq. 4-11-11 is the complete relation between e_a and θ . However, it contains the effects of two parameters that are negligible in many cases but are included for completeness. The spring load, K_{se} , is generally zero and the armature inductance, L_a , is small. If these two factors are neglected, Eq. 4-11-11 can be written as

$$e_a = \left(\frac{R_a J_e}{K_T} \right) \frac{d^2 \theta}{dt^2} + \left(\frac{R_a D_e}{K_T} + K_{av} \right) \frac{d\theta}{dt} \quad (4-11-12)$$

The Laplace transform of Eq. 4-11-12, as taken with the assumption of rest conditions, is

$$E_a(s) = s \left[\left(\frac{R_a J_e}{K_T} \right) s + \left(\frac{R_a D_e}{K_T} + K_{av} \right) \right] \Theta(s) \quad (4-11-13)$$

The desired transfer function can now be written as

$$G(s) = \frac{\Theta(s)}{E_a(s)} = \frac{K_T / R_a J_e}{s \left[s + \frac{1}{J_e} \left(D_e + \frac{K_T K_{av}}{R_a} \right) \right]} \quad (4-11-14)$$

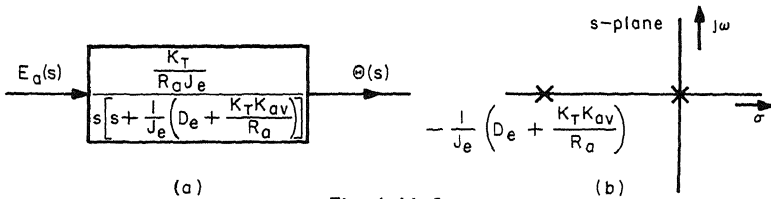


Fig. 4-11-2.

Fig. 4-11-2a shows the most commonly used block diagram representation of the armature controlled d-c servomotor. The dynamic characteristics can be represented by the s -plane plot of the poles and zeros of the transfer function of Eq. 4-11-14 as shown in Eq. 4-11-26. The motor being considered establishes a constant rate of change of armature position when a constant armature input voltage is applied. An integration in the signal path is required for this type of behavior. This characteristic is indicated by the pole at the origin in Fig. 4-11-2b.

In some cases the motor speed is considered as the output. An armature voltage to motor speed transfer function can be obtained by multiplying Eq. 4-11-14 by s . The result is equivalent to a differentiation for rest conditions. Figs. 4-11-3a and 4-11-3b show the altered forms of Figs. 4-11-2a and 4-11-2b when speed is the desired output. The fact that a constant armature voltage results in a constant motor speed means that no integration exists between input and output.

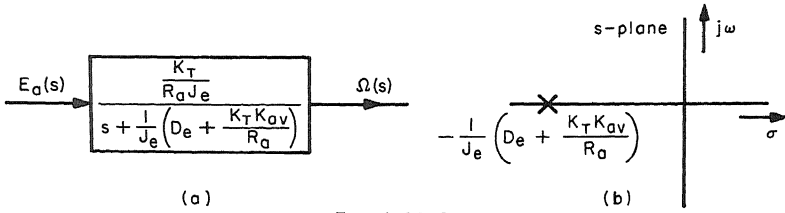


Fig. 4-11-3.

The constants K_{av} and K_T of Eqs. 4-11-7 and 4-11-8 are related by a constant dependent only on the system of units used. An easy method for finding this constant makes use of the fact that the electrical power developed by the armature current flowing through the armature counter emf, e_g , must be equal to the mechanical power as found from the developed torque, T_D . The electrical power is

$$e_g i_a = K_{av} \omega i_a \quad \text{watts} \quad (4-11-15)$$

The mechanical power is

$$T_D \omega = K_T i_a \omega \quad (\text{ft-lb})/\text{sec} \quad \text{or} \quad (\text{oz-in.})/\text{sec} \quad (4-11-16)$$

In terms of a common unit such as horsepower, Eqs. 4-11-15 and 4-11-16 should result in the same numerical value. This can be expressed as

$$\frac{K_{av} \omega i_a [\text{watts}]}{746} = \frac{K_T i_a \omega [(\text{ft-lb})/\text{sec}]}{550} \quad (4-11-17)$$

Therefore, for typical dimensional units

$$\frac{K_{av}}{K_T} = \frac{746}{550} = 1.356 \quad (\text{watt-sec})/(\text{ft-lb}) \quad (4-11-18)$$

It frequently occurs that some of the constants for the d-c motor must be obtained from torque vs. speed curves. When the motor speed, $\omega = d\theta/dt$, is constant, the armature current is constant and Eqs. 4-11-4, 4-11-7, and 4-11-8 can be combined to give

$$T_D = \frac{K_T}{R_a} (e_a - K_{av} \omega) \quad (4-11-19)$$

For fixed values of armature voltage Eq. 4-11-19 represents various torque vs. speed curves such as those shown in Fig. 4-11-4. The constants, K_{av} , K_T , and R_a are obtained from the curves of Fig. 4-11-4 as follows. For an idealized no-load condition $T_D = 0$ and from Eq. 4-11-19

$$K_{av} = \frac{e_a}{\omega_{\text{no load}}} \quad (4-11-20)$$

When the motor is stalled, the speed is zero and again from Eq. 4-11-19

$$\frac{K_T}{R_a} = \frac{T_{D \text{ stall}}}{e_a} \quad (4-11-21)$$

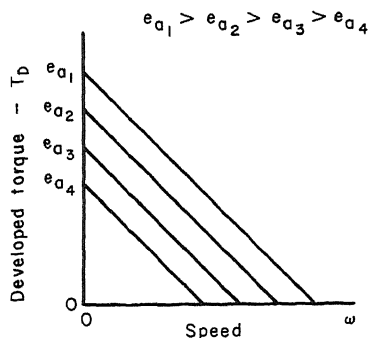


Fig. 4-11-4.

Eqs. 4-11-20 and 4-11-21 together with the K_{av}/K_T relation of Eq. 4-11-18 allow the determination K_T , K_{av} , and R_a .

As noted in Chapter 6, the slope of the torque vs. speed curves of Fig. 4-11-4 is an important factor in damping the armature controlled d-c motor. This factor is

$$\left| \frac{dT_D}{d\omega} \right| = \frac{K_T K_{av}}{R_a} \quad (4-11-22)$$

In the transfer function of Eq. 4-11-14, $K_T K_{av}/R_a$, adds to the effective viscous friction, D_e .

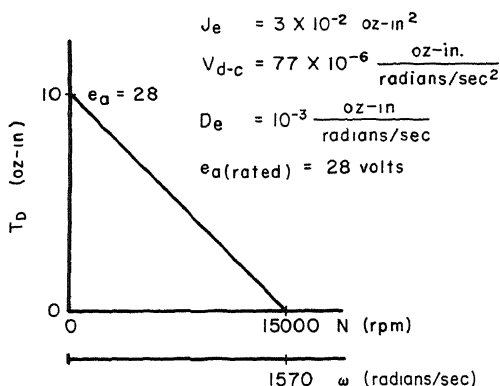


Fig. 4-11-5.

Example 4-11-1. As an example of the procedure used for obtaining an armature controlled d-c motor position transfer function, a motor with the parameters of Fig. 4-11-5 is studied. Eq. 4-11-20 can be used with the no-load speed of 15,000 rpm and the armature voltage of 28 v to find K_{av} , as follows:

$$K_{av} = \frac{28 \text{ v}}{1570 \text{ radians/sec}} = 0.0178 \text{ v/(radians/sec)} \quad (4-11-23)$$

From the stall torque at zero speed and Eq. 4-11-21

$$\frac{K_T}{R_a} = \frac{10 \text{ oz-in.}}{28 \text{ v}} = 0.357 \text{ (oz-in.)/v} \quad (4-11-24)$$

For the ounce-inch combination of units

$$\frac{K_{av}}{K_T} = 7.07 \times 10^{-3} \text{ (watt-sec)/(in.-oz)} \quad (4-11-25)$$

Eqs. 4-11-24 and 4-11-25 can be combined to give

$$K_T = \frac{0.0178 \text{ v/(radians/sec)}}{7.07 \times 10^{-3} \text{ (watt-sec)/(in.-oz)}} \quad (4-11-26)$$

$$= 2.52 \text{ (oz-in.)/amp} \quad (4-11-27)$$

$$R_a = \frac{2.52 \text{ (oz-in.)/amp}}{0.357 \text{ (oz-in.)/v}} = 7.07 \text{ ohms} \quad (4-11-28)$$

Direct substitution of numerical values into Eq. 4-11-14 now leads to the desired transfer function.

$$\frac{K_T}{R_a J_e} = \frac{0.357 \text{ (oz-in.)/v}}{77 \times 10^{-6} \text{ (oz-in.)/(radians/sec}^2\text{)}} \quad (4-11-29)$$

$$= 4.64 \times 10^3 \text{ radians/(v-sec}^2\text{)} \quad (4-11-30)$$

$$\frac{1}{J_e} \left(D_e + \frac{K_T K_{av}}{R_a} \right) = \left(\frac{1}{77 \times 10^{-6}} \right) \left[10^{-3} + (0.0178)(0.357) \right] \frac{\text{(oz-in.)/(radians/sec)}}{\text{(oz-in.)/(radians/sec}^2\text{)}} \quad (4-11-31)$$

$$= 95.8 \text{ sec}^{-1} \quad (4-11-32)$$

$$G(s) = \frac{\Theta(s)}{E_a(s)} = \frac{4.64 \times 10^{-3}}{s(s + 95.8)} \quad (4-11-33)$$

4-12. D-c Motors (Field Controlled). An alternate method of controlling d-c motors is by varying the field voltage as the input, while maintaining a constant armature current. A large current source is required in order to supply the armature current but the controlled field current is small. The basic equations which define d-c motor operation

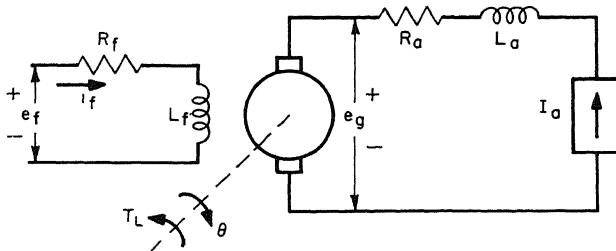


Fig. 4-12-1.

are Eqs. 4-11-1 through 4-11-6 of Art. 4-11. Fig. 4-12-1 shows the schematic of the field controlled motor. The symbols that are used in

this analysis are identical with those listed in Fig. 4-11-1 except that i_a becomes a constant, I_a , and two new constants, K'_T and K'_{av} are used. The basic motor equations of Art. 4-11 are listed below in slightly altered form.

$$i_f R_f + L_f \frac{di_f}{dt} = e_f \quad (4-12-1)$$

$$K'_{av} i_f \frac{d\theta}{dt} = e_g \quad (4-12-2)$$

$$K'_T I_a i_f = T_D \quad (4-12-3)$$

$$J_e \frac{d^2\theta}{dt^2} + D_e \frac{d\theta}{dt} + K_{se}\theta = T_L \quad (4-12-4)$$

The counter emf term, e_g , is nonlinear in this case since it involves a product of variables. However, since the armature current is held constant, the nonlinear term does not enter the transfer function for the field controlled d-c motor.

A transfer function that relates motor output position to field control voltage is the desired goal in this analysis. Since $T_D = T_L$, Eqs. 4-12-3 and 4-12-4 are equal and can be solved for i_f as follows:

$$i_f = \frac{1}{K'_T I_a} \left[J_e \frac{d^2\theta}{dt^2} + D_e \frac{d\theta}{dt} + K_{se}\theta \right] \quad (4-12-5)$$

Eq. 4-12-5 is now used to eliminate i_f in Eq. 4-12-1. After simplification, the following relation is obtained:

$$(K'_T I_a) e_f(t) = (L_f J_e) \frac{d^3\theta}{dt^3} + (L_f D_e + R_f J_e) \frac{d^2\theta}{dt^2} + (R_f D_e + L_f K_{se}) \frac{d\theta}{dt} + R_f K_{se}\theta \quad (4-12-6)$$

The spring constant K_{se} is again absent in most systems; however, the field inductance cannot be neglected as was the armature inductance in Art. 4-11. For no spring load the simplified relation is

$$\left(\frac{K'_T I_a}{L_f J_e} \right) e_f(t) = \frac{d^3\theta}{dt^3} + \left(\frac{D_e}{J_e} + \frac{R_f}{L_f} \right) \frac{d^2\theta}{dt^2} + \left(\frac{R_f D_e}{L_f J_e} \right) \frac{d\theta}{dt} \quad (4-12-7)$$

The Laplace transform of Eq. 4-12-7 as taken for forming a transfer function is

$$\left(\frac{K'_T I_a}{L_f J_e} \right) E_f(s) = s \left[s^2 + \left(\frac{D_e}{J_e} + \frac{R_f}{L_f} \right) s + \left(\frac{R_f D_e}{L_f J_e} \right) \right] \Theta(s) \quad (4-12-8)$$

The last step in forming the transfer function gives

$$G(s) = \frac{\Theta(s)}{E_f(s)} = \frac{(K'_T I_a / L_f J_e)}{s(s + D_e/J_e)(s + R_f/L_f)} \quad (4-12-9)$$

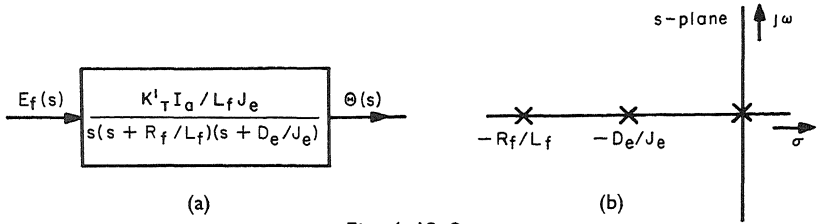


Fig. 4-12-2.

Fig. 4-12-2a shows the block diagram representation for the field controlled d-c motor when motor position is the output. The dynamic characteristics of this motor are depicted by the s -plane pole-zero plot of Fig. 4-12-2b. The integrating pole at the origin in this figure indicates that a constant field voltage results in a constant rate of change of the output position. When motor speed is the desired output, the pole at the origin is not a part of the motor transfer function.

A relation between K'_T and K'_{av} can be obtained by equating electrical and mechanical power as was done in Art. 4-11.

$$e_g I_a = K'_{av} i_f \omega I_a \quad \text{watts} \quad (4-12-10)$$

$$T_D \omega = K'_T i_f I_a \omega \quad (\text{ft-lb})/\text{sec} \quad \text{or} \quad (\text{oz-in.})/\text{sec} \quad (4-12-11)$$

$$\frac{K'_{av} i_f \omega I_a [\text{watts}]}{746} = \frac{K'_T i_f I_a \omega [(\text{ft-lb})/\text{sec}]}{550} \quad (4-12-12)$$

$$\frac{K'_{av}}{K'_T} = \frac{746}{550} \quad (\text{watt-sec})/(\text{ft-lb}) \quad (4-12-13)$$

This is the same relation as was obtained in Art. 4-11 between K_{av} and K_T . This is the reason for keeping I_a separate in the equations in this section in spite of the fact that I_a is constant.

Example 4-12-1. The transfer function for a field controlled d-c motor with the following constants is to be found. With $e_f = 100$ v and $I_a = 20$ amp, the motor speed is 1,200 rpm. The field time constant, L_f/R_f , is 0.1 sec while the mechanical time constant, J_e/D_e , is 5 sec. Since no information is given to allow direct calculation of K'_T , L_f , and J_e , some alternate scheme must be employed. The speed for a certain field voltage and armature current is stated, thus allowing an approach that is based on the motor speed. Multiplying Eq. 4-12-9 by s gives the velocity transfer function

$$\frac{\Omega(s)}{E_f(s)} = \frac{(K'_T I_a / L_f J_e)}{(s + R_f/L_f)(s + D_e/J_e)} \quad (4-12-14)$$

If a constant field voltage, E_f , is applied, the speed is

$$\Omega(s) = \frac{(K'_T I_a / L_f J_e) E_f}{s(s + R_f/L_f)(s + D_e/J_e)} \quad (4-12-15)$$

The steady angular velocity, as found by application of the final-value theorem of Chapter 2, is

$$\omega_{ss}(t) = \lim_{s \rightarrow 0} s\Omega(s) = \frac{(E_f)(K'_T I_g / L_f J_e)}{(R_f / L_f)(D_e / J_e)} \quad (4-12-16)$$

For the given values of E_f , ω_{ss} , R_f / L_f , D_e / J_e Eq. 4-12-16 is

$$\frac{1200}{60} \times 2\pi = \frac{(100)(K'_T I_a / L_f J_e)}{(10)(0.2)} \quad (4-12-17)$$

$$\frac{K'_T I_a}{L_f J_e} = 2.51 \text{ radians/v-sec}^3 \quad (4-12-18)$$

The transfer function relating motor position to field voltage can now be written as

$$\frac{\Theta(s)}{E_f(s)} = \frac{2.51}{s(s + 10)(s + 0.2)} \quad (4-12-19)$$

4-13. Two-Phase A-c Servomotors. Frequently a small a-c servomotor is used to rotate a load in a position or velocity control system. These motors have two basic field windings located 90° apart in space and excited by a-c voltages that are 90 electrical degrees apart in time. This arrangement results in a rotating magnetic flux field that induces a voltage in a closed armature circuit. A circulating armature current results from the induced voltage and this current establishes its own magnetic field. The reaction between the rotating magnetic flux due to the field windings and the magnetic flux due to the armature current causes a torque. This torque causes rotation of the armature.

One of the two field windings of an a-c servomotor is excited by a fixed amplitude a-c line voltage. The amplitude of the other field voltage is used as the motor input in order to control the output torque. The first field referred to above is called the fixed field or reference field. The second field is called the control field. Since normal operation of the two-phase servomotor requires unequal field voltage amplitudes, the operation is said to involve unbalanced voltages. This complicates the analysis of this device. It is possible to obtain the following approximate relation by using special methods.⁶

$$T_D = T_s \frac{e_{cf}}{E_{ff}} - \frac{K_1}{2} \left[1 + \left(\frac{e_{cf}}{E_{ff}} \right)^2 \right] \frac{d\theta}{dt} \quad (4-13-1)$$

Fig. 4-13-1. shows the schematic of the two-phase a-c servomotor, and the following symbols are used in connection with this figure:

E_{ff} = amplitude of the fixed field a-c voltage

e_{cf} = time-varying amplitude (envelope) of the control field voltage

T_s = stall torque with balanced rated phase voltages applied

⁶ John E. Gibson and Franz B. Tuteur, *Control System Components*, McGraw-Hill Book Co., Inc., New York, 1958, Chapter 7.

K_1 = magnitude of the torque-speed curve at zero speed with rated balanced voltages applied

θ = armature position

T_D = developed torque

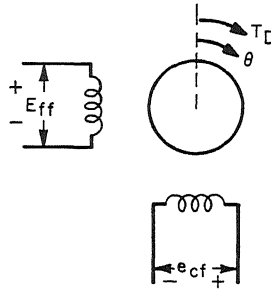


Fig. 4-13-1.

Eq. 4-13-1 involves only the envelope or time-varying amplitude of the control field voltage, and not the actual instantaneous voltage. This situation is similar to that encountered in Art. 4-10 which discusses synchros. The actual signal information is contained in the envelope

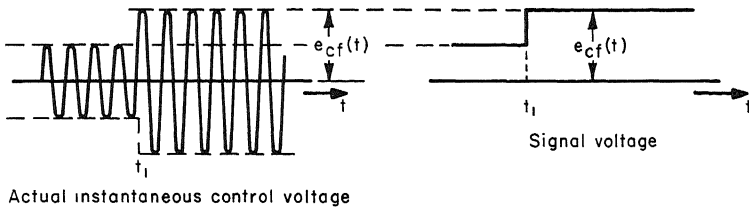


Fig. 4-13-2.

as shown in Fig. 4-13-2. Eq. 4-13-1 also neglects the electrical time constants of the field and armature circuits. In spite of the approximations that have been used, Eq. 4-13-1 is still nonlinear because of the $(K_1/2)(e_{cf}/E_{ff})^2 d\theta/dt$ term. This term must be neglected if a linear equation is desired. Since normal positioning control systems require the motor to work with very small control field voltages, the ratio $(e_{cf}/E_{ff})^2$ is generally small compared to one. Therefore the error caused by neglecting $(K_1/2)(e_{cf}/E_{ff})^2 d\theta/dt$ in Eq. 4-13-1 is not too serious. The linearized torque equation can now be written.

$$T_D \approx \frac{T_s}{E_{ff}} e_{cf} - \frac{K_1}{2} \frac{d\theta}{dt} \quad (4-13-2)$$

A transfer function which relates a-c servomotor position to control field voltage amplitude can now be found if a particular load is assumed. Such a transfer function is commonly called an envelope transfer function.

For a load made up of an effective moment of inertia, J_e , and effective viscous friction, D_e , the load torque, T_L , is

$$T_L = J_e \frac{d^2\theta}{dt^2} + D_e \frac{d\theta}{dt} \quad (4-13-3)$$

The developed torque of Eq. 4-13-2 must equal the load torque; therefore,

$$J_e \frac{d^2\theta}{dt^2} + D_e \frac{d\theta}{dt} = \frac{T_s}{E_{ff}} e_{cf} - \frac{K_1}{2} \frac{d\theta}{dt} \quad (4-13-4)$$

For zero initial conditions the Laplace transform of Eq. 4-13-4 can be arranged as follows,

$$s[J_e s + (D_e + \frac{K_1}{2})]\Theta(s) = \frac{T_s}{E_{ff}} E_{cf}(s) \quad (4-13-5)$$

The desired envelope transfer function is

$$G(s) = \frac{\Theta(s)}{E_{cf}(s)} = \frac{T_s/J_e E_{ff}}{s[s + \frac{1}{J_e}(D_e + \frac{K_1}{2})]} \quad (4-13-6)$$

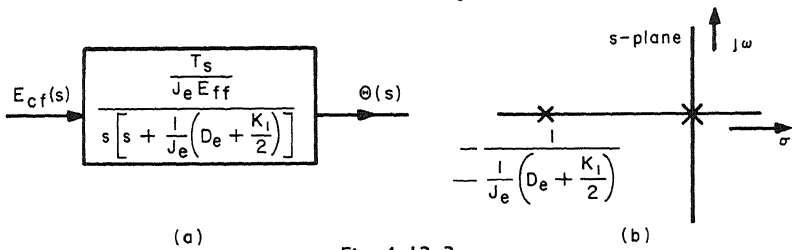


Fig. 4-13-3.

Fig. 4-13-3a shows the block diagram representation of the a-c servomotor while Fig. 4-13-3b shows its s -plane pole-zero plot. The similarity between the pole-zero plot for the armature controlled d-c motor in Fig. 4-11-26 and for the a-c servomotor in Fig. 4-13-3b should be noted. Both pole-zero plots contain an integrating pole at the origin because a constant input voltage causes a constant rate of change of the output position. In addition both pole-zero plots have a negative real pole whose location is partly determined by the slope of the torque v. speed curves.

Example 4-13-1. The envelope transfer function for a two-phase, 400 cps, a-c servomotor is to be found. The servomotor characteristics are given in Fig. 4-13-4. At stall $T_s = 1.0$ oz.-in. while at zero speed the magnitude of the slope of the torque vs. speed curve is

$$K_1 = \frac{1 \text{ oz.-in.}}{8000 \text{ rev/min}} \times \frac{60 \text{ sec/min}}{2\pi \text{ radians/rev}} \quad (4-13-7)$$

$$= 1.19 \times 10^{-3} \text{ oz.-in.-sec} \quad (4-13-8)$$

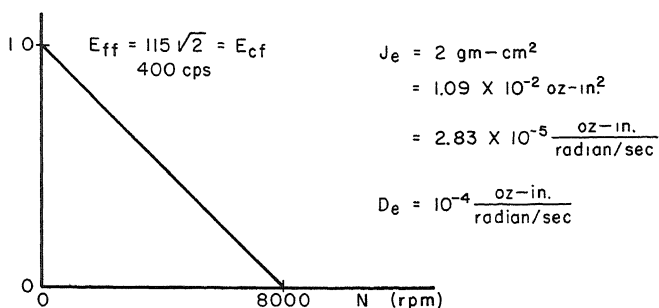


Fig. 4-13-4.

From the above data

$$\frac{T_s}{J_e E_{ff}} = \frac{1}{(2.83 \times 10^{-5})(115 \sqrt{2})} \frac{\text{oz-in.}}{[(\text{oz-in.})/(\text{radian/sec}^2)]v} \quad (4-13-9)$$

$$= 218 \text{ radians}/(\text{sec}^2 \cdot v.) \quad (4-13-10)$$

$$\frac{1}{J_e} \left[D_e + \frac{K_1}{2} \right] = \frac{1}{2.83 \times 10^{-5}} [10^{-4} + 5.97 \times 10^{-4}] \frac{(\text{oz-in.})/(\text{radian/sec})}{(\text{oz-in.})/(\text{radian/sec}^2)} \quad (4-13-11)$$

$$= 24.6 \text{ sec}^{-1} \quad (4-13-12)$$

The desired transfer function can now be written as

$$G(s) = \frac{\Theta(s)}{E_{ef}(s)} = \frac{218}{s[s + 24.6]} \quad (4-13-13)$$

PROBLEMS

4-1. Find an equivalent spring constant for the system of springs shown in Fig. 4-P-1 and determine the potential energy stored in the springs when $x = 2$ in., if the springs are unstretched at $x = 0$.

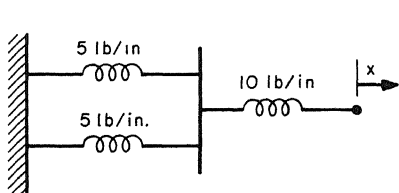


Fig. 4-P-1.

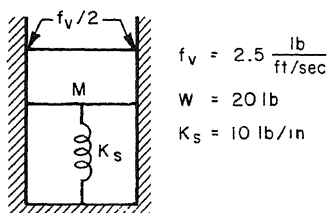


Fig. 4-P-2.

4-2. The mass in Fig. 4-P-2 is initially pushed down 2 in. from its rest position and then released. The mass is firmly attached to the spring. Find the equation of motion for the mass.

4-3. Fig. 4-P-3 shows a mechanical system that is initially at rest. The force, $F(t) = 10$ lbs, is applied to mass M_1 as shown at $t = 0$. Solve the system for $X_2(s)$

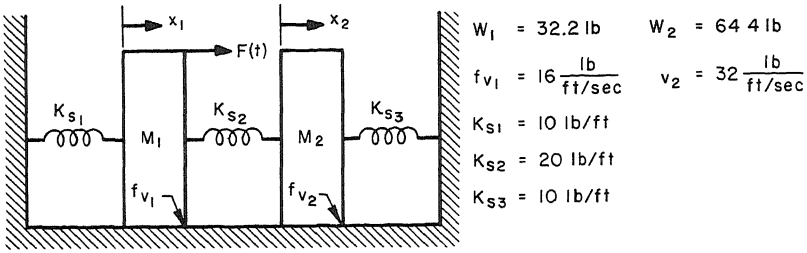


Fig. 4-P-3.

and then use the final-value theorem of Chapter 2 to find the steady-state position of mass M_2 .

4-4. The mechanical system of Fig. 4-P-4 has $x_2 = x_2(0)$ in. when a sinusoidal force, $F(t) = F_m \sin \omega t$ is applied. Find the steady-state transfer function $X_2(j\omega)/F(j\omega)$. What limitations exist on the validity of the system equations?

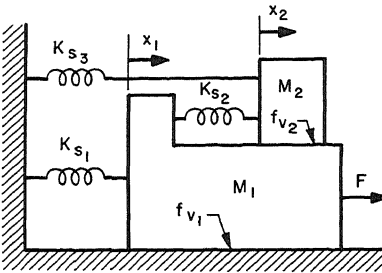


Fig. 4-P-4.

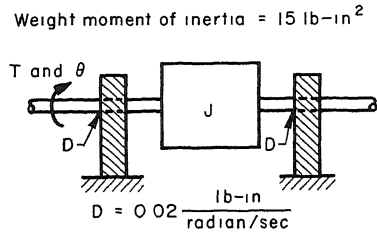


Fig. 4-P-5.

4-5. Fig. 4-P-5 shows a rotational system in which viscous friction exists at the two bearings. A torque of 1 lb-in. is applied at $t = 0$ when the body is at rest.

(a). Find the equation for the angular velocity of the body. Show that the response due to a unit impulse of torque is the derivative of the response due to the unit-step function applied torque.

(b) For a unit-step applied torque find the power dissipated as heat at the bearings after steady state has been reached. What kinetic energy is stored in the rotating body?

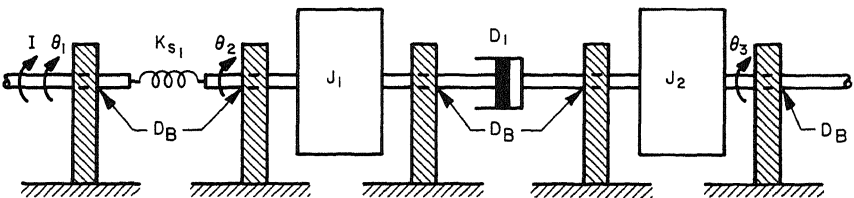


Fig. 4-P-6.

4-6. Fig. 4-P-6 shows a rotational system which is initially at rest. Viscous friction at all bearings as well as at the coupling between bodies J_1 and J_2 . Write the system differential equations and find the transfer function $\Theta_3(s)/T(s)$.

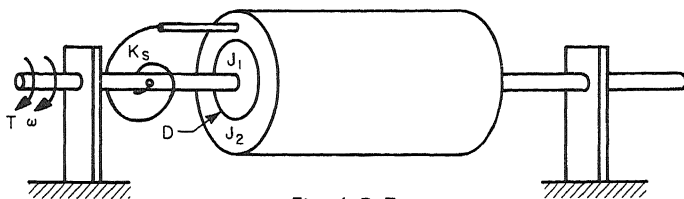


Fig. 4-P-7.

4-7. Find the angular velocity equation for body J_2 in Fig. 4-P-7, if body J_1 is suddenly given a velocity of ω radians/sec. The system is initially at rest. What form (step function, impulses, doublets, etc.) of torque is needed to set body J_1 in motion as stated above.

4-8. Find the effective inertia at shaft θ_1 and determine the angular acceleration of bodies J_2 and J_3 in Fig. 4-P-8 if a torque $T = 10$ oz-in. is applied. The gears have negligible inertia and friction. Bodies J_2 and J_3 are each aluminium cylinders 2 in. long and 4 in. in diameter.

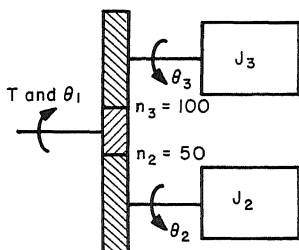


Fig. 4-P-8.

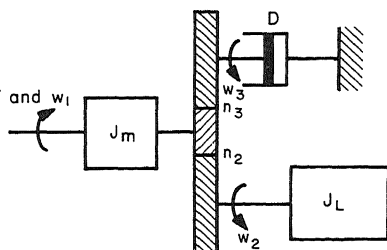


Fig. 4-P-9.

4-9. Find the transfer functions $\Omega_3(s)/T(s)$ and $\Omega_2(s)/T(s)$ for the rotational system of Fig. 4-P-9. Draw a block diagram representation of this system showing both Ω_2 and Ω_3 as outputs and T as the input. Draw an s -plane pole-zero plot for each transfer function above and describe the natural response modes of this system. Do decaying exponentials, constants, etc., result from impulse excitation? What average power is dissipated as heat in D in the steady state when a sinusoidal torque, $T = T_m \sin \omega t$ is applied?

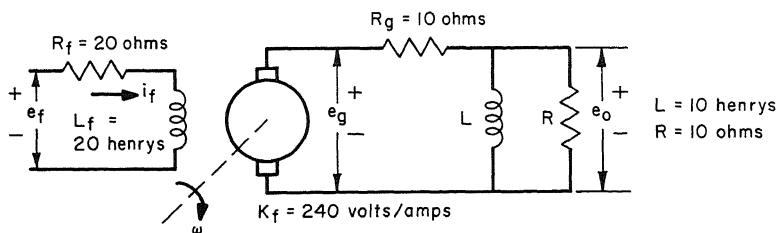


Fig. 4-P-10.

4-10. (a) The field of an ordinary d-c generator has $R_f = 120$ ohms, $L_f = 20$ h, and $K_f = 240$ v/amp (ohms). Draw a block diagram of this generator with open-circuit generated voltage, e_g , as the output and field voltage, e_f , as the input. Sketch an s -plane pole-zero pattern for this unloaded generator.

(b) The generator of part (a) is loaded by the circuit of Fig. 4-P-10. The generator has an armature resistance of 10 ohms and negligible inductance. Determine the block diagram that relates the open circuit generated voltage to the output voltage, e_o . Plot an over-all pole-zero pattern.

4-11. A step function field voltage $e_f = 120$ v is suddenly applied to the generator of Prob. 4-10b. Plot the voltage $e_o(t)$ and relate its characteristics to the pole-zero plot of Prob. 4-10b.

4-12. In steady-state operation an amplidyne with rated control field voltage of 28 vd-c applied provides an open-circuit output voltage of 350 v. The control field time constant, L_{cf}/R_{cf} , is 0.1 sec while the quadrature field time constant is 0.2 sec. Determine the transfer function which relates open-circuit output to control field voltage. Sketch the s -plane pole-zero pattern for this amplidyne and describe the natural response modes (impulse response terms).

4-13. An armature controlled d-c motor has an effective load moment of inertia of 9×10^{-3} lb-in². and negligible load viscous friction. The no-load speed with rated voltage applied is 12,000 rpm while the other rated characteristics are:

Speed = 10,000 rpm

Torque = 0.05 lb-in.

Armature voltage = 6 v.

Assume a straight line torque vs. speed curve. Find the transfer function which relates armature speed to armature voltage.

4-14. The d-c motor of Ex. 4-11-1 has rated voltage applied. Find the steady-state angular velocity by using the final-value theorem. Plot a steady-state load torque curve ($T = D\omega$) on a graph of motor torque v. speed and determine the steady-state speed from this graph.

4-15. A synchro generator or transmitter is connected to a control transformer as shown in Fig. 4-10-3. The rotor of the control transformer is held fixed at $\theta_0 = 0^\circ$ while the generator shaft is turned at a constant angular velocity, ω_s . The supply line frequency is ω_c . Write an expression for the actual instantaneous voltage between the control transformer rotor terminals if the generator rotor angle, θ_q , is zero at $t = 0$. Determine the frequencies present in this voltage. What has happened to the original frequency ω_s ?

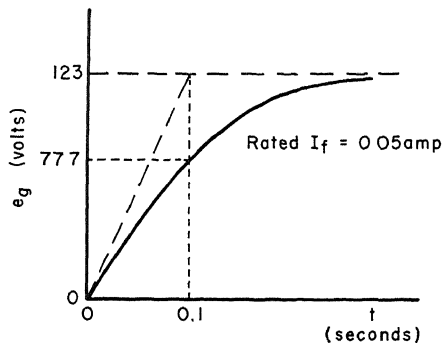


Fig. 4-P-16.

4-16. An amplidyne with a rated output voltage of 115 V at a load current of 0.4 amp is driven by a current source. A step input current of 0.05 amp causes the open-circuit response shown in Fig. 4-P-16. Determine the transfer function

which relates open-circuit output voltage to input control field current. Find the resistance in the output circuit.

4-17. The amplidyne of Prob. 4-16 is used to drive an armature controlled d-c motor. The inductance of the motor armature and the amplidyne output circuit can be neglected. In steady state an amplidyne field current of 0.05 amp causes a motor speed of 16,000 rpm with a load torque of 0.1 lb-in. The motor delivers a stall torque of 0.5 lb-in. for the 0.05 amp amplidyne input. If a straight line torque vs. speed curve is assumed for the motor, can the motor transfer function be found? If so, find it; if not, find as much of it as possible and determine an additional test to complete the transfer function.

4-18. Plot the response of the field control d-c motor of Ex. 4-12-1 to a unit impulse. What property of the motor is indicated by constant steady-state position with no input? What would be changed in the motor response if a current source were used to provide the field input?

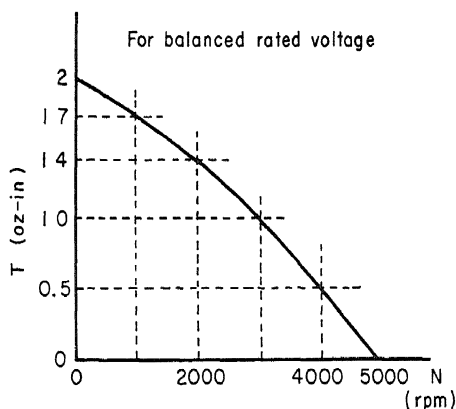


Fig. 4-P-19.

4-19. A 115 v rms, 400 cps two-phase a-c servomotor has the rated torque vs. speed curve of Fig. 4-P-19. The rotor moment of inertia is 3 gm-cm^2 . An additional load $J_L = 2 \times 10^{-2} \text{ ozin.}^2$ and $D_L = 4 \times 10^{-3} \text{ oz-in.-sec}$ is coupled to the motor. Find the envelope transfer function which relates motor speed to control field voltage. Assume a 90° phase angle between field voltages.

4-20. If the control field voltage for the motor of Prob. 4-19 has a sinusoidal envelope $e_{cf}(t) = 115\sqrt{2} \sin 60t$ and if rated fixed field voltage is applied, find the motor speed in the steady state.

HYDRAULIC AND PNEUMATIC SYSTEMS

5-1. Introduction. Many control systems have components that use hydraulic and pneumatic principles. Certain advantages may be gained by using these components whenever the cost and complexity of a hydraulic fluid or an air supply are justified. For high power actuators, hydraulic components give a larger output power to size ratio than do electrical components. Air systems have three principal advantages. First, shock waves are not transmitted; second, the used air can generally be exhausted to the atmosphere; third, there is little danger of fire or explosion.

This chapter discusses the various parameters and variables of hydraulic and pneumatic components and develops approximate transfer functions for a few such devices. The fact that the transfer functions are based on linear operation greatly reduces their accuracy. This is due to the fact that the basic flow relations are nonlinear. A really accurate study of hydraulic and pneumatic devices would require either a computer set-up which accounts for the various nonlinear flow relations, saturation effects, and limits on variables or an actual physical mock-up of the device being studied.

5-2. Hydraulic and Pneumatic System Parameters and Variables. Most of the basic hydraulic system parameters seem to fall into a different category from that of the mechanical and electrical parameters. No particular symbols represent such items as density, viscosity, compressibility, specific heat, specific gravity, specific weight, and system dimensions which are in reality the parameters of hydraulic and pneumatic systems. Generally a schematic of some kind shows the system layout and indicates the system variables—that is, pressures (heads), volume or weight flow rates, and temperatures plus possibly some system dimensions. Most of the above parameters and variables need no discussion since they are introduced in basic physics. However the parameters, viscosity and compressibility, are reviewed in the following paragraphs.

The term viscosity is used to represent the resistance-to-flow property of a fluid. When adjacent layers of a fluid move with different velocities, a shearing stress exists between these layers because of the molecular

properties of the fluid. The shearing rate for two surfaces of area A and separation l when moving with a relative velocity v is v/l . If a shearing force F is required to establish this relative motion in a fluid, the shearing stress is F/A . For fluids the ratio of shearing stress to shearing rate is called the absolute viscosity μ .

$$\mu = \frac{F/A}{v/l} \quad (5-2-1)$$

The absolute viscosity is also a function of temperature and pressure. In the English system of units the dimensions of μ are pound-seconds/foot² or ounce-seconds/inch². In the cgs system the dimensions of μ are dyne-seconds/centimeters², termed a poise. In some cases a kinematic viscosity is used. It is equal to the absolute viscosity, μ , divided by the fluid density. The kinematic viscosity has dimensions such as feet²/second or centimeters²/second, termed a stoke. Viscosity appears again in the flow equations.

While liquids are normally considered incompressible, the high pressures of some hydraulic systems make the compressibility of the fluid an important factor. A positive increment in pressure Δ_p in a fluid causes a negative increment in volume per unit volume that is proportional to the increase in pressure. A term called the bulk modulus, B , is used to account for this effect, thus

$$B = - \lim_{\Delta p \rightarrow 0} \frac{\Delta_p}{\Delta \text{vol}/\text{vol}} \quad (5-2-2)$$

A typical value of B for hydraulic fluids is 2.5×10^5 psi.

The auxiliary concept of laminar and nonlaminar flow in hydraulic systems determines the flow equation that applies. At relatively low velocities, fluid flow occurs in well-defined streamlines with no turbulence or swirling motion. This is called laminar flow. A critical region of velocity and pressure is reached in which the flow changes to nonlaminar or turbulent flow. A dimensionless factor called the Reynolds number is used to determine the type of flow. Reynolds numbers from about 2,400 to 3,000 generally denote the beginning of turbulent flow. For circular pipe the Reynolds number is defined as

$$\text{Reynolds number} = \frac{(\text{Average velocity})(\text{Pipe diameter})}{\text{Kinematic viscosity}} \quad (5-2-3)$$

5-3. Fluid Flow Laws. The first relation that is needed in the study of fluid flow is an equation for nonlaminar flow through restrictions and orifices. The volume flow rate q , for such conditions is

$$q = C_d A \sqrt{2g(h_1 - h_2)} \quad (5-3-1)$$

in which A is the cross-sectional area of the restriction, g is the acceleration of gravity, $h_1 - h_2$ is the drop in head, and C_d is a discharge coefficient.

Eq. 5-3-1 is a nonlinear relation between the variable, drop in head and volume flow rate.

For laminar flow in pipes the volume flow rate equation takes the form given by Eq. 5-3-2.

$$q = \frac{\pi D^4 (h_1 - h_2) \rho}{128 \mu L} \quad (5-3-2)$$

In this equation D is the pipe diameter, L is pipe length, ρ is fluid specific weight, and μ is the absolute viscosity of the fluid. As long as the Reynolds number is below the range given in Art. 5-2, Eq. 5-3-2 can be used to calculate volume flow rates.

Before proceeding to various applications, hydraulic work and energy relations are reviewed. Since work is equal to force times linear displacement or torque times angular displacement, work in hydraulic systems is given by

$$W = (pA)(\text{Distance}) = p(\text{Volume}) \quad (5-3-3)$$

Constant pressure is assumed in Eq. 5-3-3 so that integration of increments is not required. Power is found for such a case as

$$\text{Power} = \frac{\text{Work}}{\text{Time}} = p \frac{(\text{Volume})}{t} = pq \quad (5-3-4)$$

The average torque output of a hydraulic motor can be found by considering that the work done in one revolution of the motor is

$$2\pi T = \text{Work done in one revolution} \quad (5-3-5)$$

$$= (\text{Drop in pressure})(\text{Volume displaced in one revolution}) \quad (5-3-6)$$

$$T = \frac{(\Delta p)(\text{Displacement in one revolution})}{2\pi} \quad (5-3-7)$$

5-4. Hydraulic Resistance and Capacitance.¹ Two auxiliary hydraulic parameters are useful in analyzing general fluid handling systems. The first of these parameters is hydraulic resistance, which is defined as

$$R_h = \frac{d(h_1 - h_2)}{dq} \quad (5-4-1)$$

Hydraulic resistance thus relates increments in flow rate to increments in the drop in head for flow through an orifice or restriction. The hydraulic resistance, R_h , is the slope of the drop in head vs. flow curve. For flow Eq. 5-3-1

$$\frac{1}{R_h} = \frac{dq}{d(h_1 - h_2)} = \frac{C_a A \sqrt{2g}}{2(h_1 - h_2)^{1/2}} = \frac{q}{2(h_1 - h_2)} \quad (5-4-2)$$

¹ Donald P. Eckman, *Automatic Process Control*, John Wiley & Sons, Inc., New York, 1958, Chapter 2.

Generally R_h varies as q and $h_1 - h_2$ changes. If R_h is evaluated at the normal operating values of flow rate and drop in head, it may be considered to be constant for relatively small increments in q and $h_1 - h_2$. An approximate linear equation can then be written for flow rate increments, q_Δ , and drop in head increments, $h_{1\Delta} - h_{2\Delta}$.

$$h_{1\Delta} - h_{2\Delta} = R_h q_\Delta \quad (5-4-3)$$

Total values of q and $(h_1 - h_2)$ are found by adding the increments to the normal operating values. If normal operation occurs at zero flow rate and zero drop in head for an orifice, R_h is zero. This result is useless and some additional knowledge of the maximum expected flow rate is needed in order to determine an average hydraulic resistance throughout the

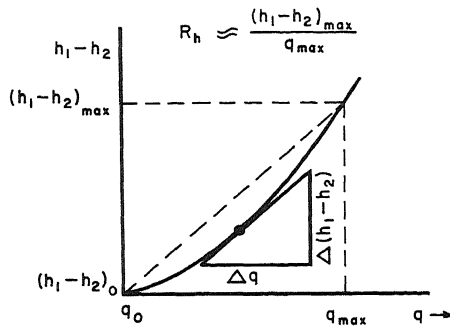


Fig. 5-4-1.

operating range. Fig. 5-4-1 illustrates a suitable choice of R_h for such a case. For laminar flow through pipes the hydraulic resistance is found from Eq. 5-3-2.

$$R_h = \frac{128\mu L}{\pi D^4 q} = \frac{h_1 - h_2}{q} \quad (5-4-4)$$

This resistance is constant since the laminar flow equation is linear.

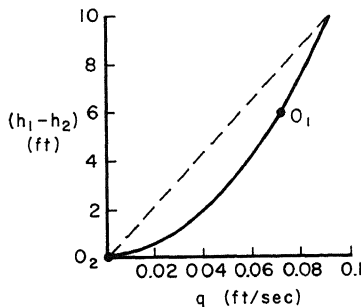


Fig. 5-4-2.

Example 5-4-1. Fig. 5-4-2 shows a flow vs. drop in head curve for an orifice with a discharge coefficient of 0.7 and a cross sectional area of 0.75 in.². The hydraulic resistance of this orifice is to be found first for an operating point O_1 of 6 ft. differential head and second for O_2 of zero head but with maximum expected variations up to 10 ft differential head. For O_1

$$R_h = \frac{2(h_1 - h_2)}{q} = \frac{2 \times 6}{0.0718} \frac{\text{ft}}{\text{ft}^3/\text{sec}} \quad (5-4-5)$$

$$= 167 \text{ sec}/\text{ft}^2 \quad (5-4-6)$$

For O_2 a suitable average R_h is found as

$$R_h \approx \frac{(h_1 - h_2)_{\max}}{q_{\max}} = \frac{10}{0.0928} = 108 \text{ sec}/\text{ft}^2 \quad (5-4-7)$$

This is actually the incremental resistance at an operating point of $h_1 - h_2 = 2.5$ ft.

The second auxiliary parameter, hydraulic capacitance C_h , is defined for tanks as

$$C_h = \frac{d(\text{Volume of fluid})}{dh} \quad (5-4-8)$$

Thus C_h relates increments in fluid volume to increments in head in a tank. For tanks of constant cross-sectional area C_h is the cross sectional area of the tank.

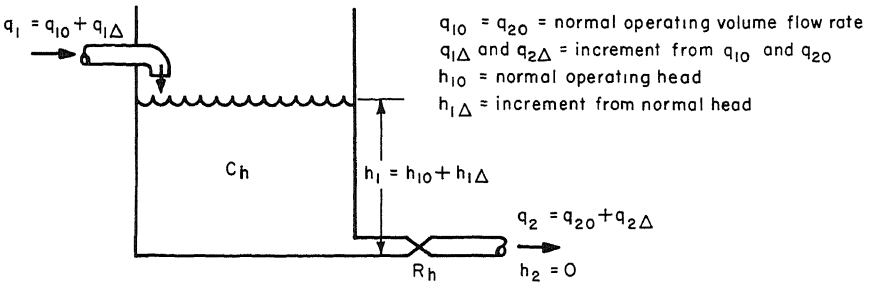


Fig. 5-4-3.

Example 5-4-2. As an example of the use of the concepts of incremental hydraulic resistance and capacitance, the fluid-handling system of Fig. 5-4-3 is analyzed to obtain a transfer function between $h_{1\Delta}$ and $q_{1\Delta}$. From volume considerations and the definition of C_h

$$q_1 - q_2 = \frac{d(\text{Volume})}{dt} = \frac{d(\text{Volume})}{dh_1} \frac{dh_1}{dt} \quad (5-4-9)$$

For $q_{10} = q_{20}$ and for h_{10} equal to a constant

$$q_{1\Delta} - q_{2\Delta} = C_h \frac{dh_{1\Delta}}{dt} \quad (5-4-10)$$

The hydraulic resistance, R_h , is used as follows

$$h_{1\Delta} - 0 = R_h q_{2\Delta} \quad (5-4-11)$$

Eqs. 5-4-10 and 5-4-11 can be combined to give

$$C_h \frac{dh_{1\Delta}}{dt} + \frac{h_{1\Delta}}{R_h} = q_{1\Delta} \quad (5-4-12)$$

The Laplace transform of Eq. 5-4-12 for zero initial conditions,

$[q_{1\Delta}(0) = 0, h_{1\Delta}(0) = 0]$, is

$$\left(C_h s + \frac{1}{R_h}\right) H_{1\Delta}(s) = Q_{1\Delta}(s) \quad (5-4-13)$$

The desired transfer function is

$$G(s) = \frac{H_{1\Delta}(s)}{Q_{1\Delta}(s)} = \frac{1/C_h}{\left(s + \frac{1}{R_h C_h}\right)} \quad (5-4-14)$$

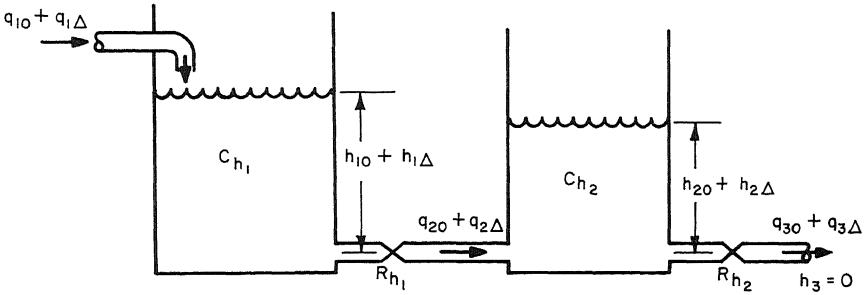


Fig. 5-4-4.

Example 5-4-3. Fig. 5-4-4 shows a second fluid-handling system. This system is to be analyzed to obtain the transfer function $H_{2\Delta}(s)/Q_{1\Delta}(s)$.

For the first tank

$$q_{1\Delta} - q_{2\Delta} = C_{h_1} \frac{dh_{1\Delta}}{dt} \quad (5-4-15)$$

The following relation can be written at restriction R_{h_1}

$$h_{1\Delta} - h_{2\Delta} = R_{h_1} q_{2\Delta} \quad (5-4-16)$$

For the second tank and restriction, the equations are

$$q_{2\Delta} - q_{3\Delta} = C_{h_2} \frac{dh_{2\Delta}}{dt} \quad (5-4-17)$$

$$h_{2\Delta} - 0 = R_{h_2} q_{3\Delta} \quad (5-4-18)$$

Eqs. 5-4-15 through 5-4-18 can be combined to give

$$R_{h_1} R_{h_2} C_{h_1} C_{h_2} \frac{d^2 h_{2\Delta}}{dt^2} + (C_{h_2} R_{h_2} + R_{h_1} C_{h_1} + C_{h_1} R_{h_2}) \frac{dh_{2\Delta}}{dt} + h_{2\Delta} = R_{h_2} q_{1\Delta} \quad (5-4-19)$$

The desired transfer function is now written by inspection.

$$G(s) = \frac{H_{2\Delta}(s)}{Q_{1\Delta}(s)} = \frac{R_{h_2}}{R_{h_1}R_{h_2}C_{h_1}C_{h_2}s^2 + (R_{h_1}C_{h_1} + R_{h_2}C_{h_2} + C_{h_1}R_{h_1})s + 1} \quad (5-4-20)$$

5-5. Hydraulic Valves and Actuators. Various specific hydraulic components are introduced in this article and linearized equations are developed to describe their performance. Since such components operate at higher pressures and with smaller volumes than the general systems of Art. 5-4, hydraulic valves, pistons, motors, and pumps require different analysis methods. These methods are introduced in the examples that follow.

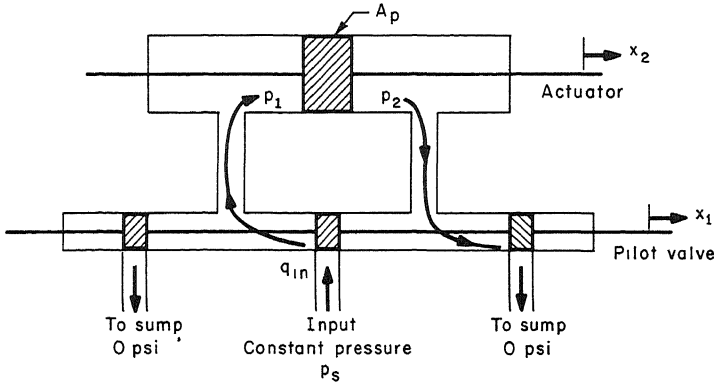


Fig. 5-5-1.

Example 5-5-1. The pilot-valve-operated linear actuator illustrated in Fig. 5-5-1 is analyzed to obtain a very approximate transfer function in this example. Movement of the pilot valve admits high pressure fluid to one side of the actuator piston. The other side of the actuator piston is simultaneously opened to the low pressure reservoir. The various pilot valve openings are orifices and from Art. 5-3 the volume flow rate, q , through an orifice is

$$q = C_d A \sqrt{\frac{2g(p_1 - p_2)}{\rho}} \quad (5-5-1)$$

Ideally the pilot valve orifice openings are proportional to the displacement x_1 . Also as a first approximation, the pressure drops $p_s - p_1$ and $p_2 - 0$ are considered constant and equal. Finally, if leakage and compression effects are neglected, the input and output flow rates are proportional to x_1 .

$$q_{in} = K_2 x_1 \quad (5-5-2)$$

This input flow must move the actuator piston with a certain velocity as indicated in Eq. 5-5-3 in which A_p is actuator piston area.

$$q_{in} = A_p \frac{dx_2}{dt} \quad (5-5-3)$$

Eqs. 5-5-2 and 5-5-3 can be combined to give

$$x_1 = \frac{A_p}{K_2} \frac{dx_2}{dt} \quad (5-5-4)$$

and the desired transfer function is

$$\frac{X_2(s)}{X_1(s)} = \frac{K_2/A_p}{s} \quad (5-5-5)$$

The velocity of the actuator piston depends on the input displacement. If the x_2 position is the output of the system that is being studied, an integration takes place between input and output. This is indicated in Eq. 5-5-5 by the pole at the origin of the Laplace s -plane. The pilot-valve-controlled piston therefore behaves somewhat similarly to the armature controlled d-c motor and the a-c servomotor of Chapter 4 in that they all give an integration between the input and output positions.

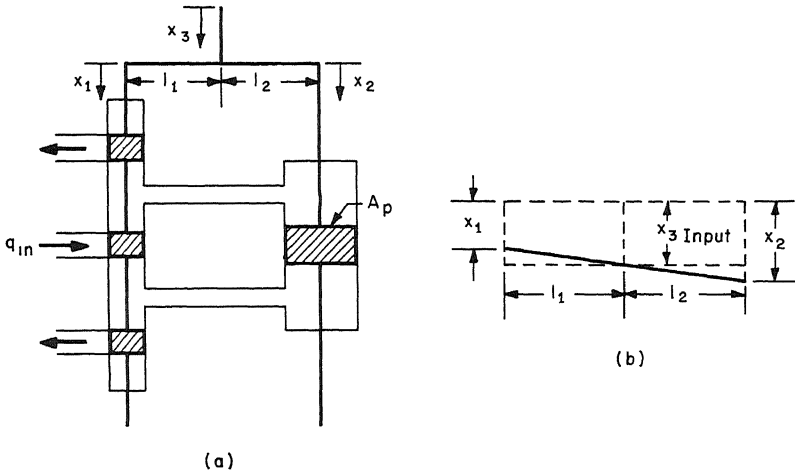


Fig. 5-5-2.

Example 5-5-2. In some cases a mechanical feedback link is added to the system of Fig. 5-5-1. This link is shown in Fig. 5-5-2a and b which illustrate the second hydraulic valve-piston example. A transfer function relating the output position x_2 to input position x_3 is to be found. The assumptions concerning fluid flow which are made in Ex. 5-5-1 are repeated here. Only small movements are considered. Eq. 5-5-4 is still valid. However x_1 , x_2 , and x_3 are all related now according to Eqs. 5-5-6 and 5-5-7 which can be written from Fig. 5-5-2b if only small movements are considered.

$$x_1 = x_3 - \left(\frac{l_1}{l_2}\right)(x_2 - x_3) \quad (5-5-6)$$

$$x_1 = \left(\frac{l_1 + l_2}{l_2}\right)x_3 - \left(\frac{l_1}{l_2}\right)x_2 \quad (5-5-7)$$

The differential equation relating input and output displacements is obtained by eliminating x_1 between Eq. 5-5-4 and Eq. 5-5-7.

$$\left(\frac{l_1 + l_2}{l_2}\right) x_3 = \frac{A_p}{K_2} \frac{dx_2}{dt} + \frac{l_1}{l_2} x_2 \quad (5-5-8)$$

The desired transfer function can now be written by inspection.

$$\frac{X_2(s)}{X_3(s)} = \frac{(K_2/A_p)[(l_1 + l_2)/l_2]}{s + K_2 l_1/A_p l_2} \quad (5-5-9)$$

The output is no longer the integral of the input. Instead a single time constant response is characteristic of the system. The integration is lost because the movement of the actuator piston cancels out the initial movement of the pilot valve caused by a movement of the input x_3 . Flow to the actuator is thus stopped and the actuator stops. Further investigation of this system is left as Prob. 5-4.

Example 5-5-3. A more detailed analysis of the valve operated actuator of Fig. 5-5-1 is made in this example. While the assumption of negligible compressibility used in Ex. 5-5-1 may be valid because of the small volume of fluid under compression, the other assumptions made in that example lead to appreciable errors. The input flow rate should be related to the drop in pressure across the input valve opening. Eq. 5-5-10

$$q_{in} = K_3 x_1 \sqrt{p_s - p_1} \quad (5-5-10)$$

gives this relationship if it is assumed that the valve opening is proportional to valve displacement, x_1 . The supply pressure, p_s , is constant while the pressure p_1 on the high pressure side of the piston is a variable. Eq. 5-5-10 is nonlinear since both a product of variables and a square root are involved. Nevertheless this does not make further approximate linear analysis impossible. There must be a relation between the load force, F_L , the actuator piston area, A_p , and the drop in pressure, $(p_1 - p_2) = p_L$, across the actuator piston. This relation is given in Eq. 5-5-11 as

$$F_L = A_p(p_1 - p_2) = A_p p_L \quad (5-5-11)$$

It is also possible to relate the valve orifice pressure drops, $(p_s - p_1)$ and $(p_2 - 0)$, to the load pressure drop, $p_1 - p_2 = p_L$, and the supply pressure, p_s .

$$\text{Total orifice drops} = p_s - p_L \quad (5-5-12)$$

With p_s fixed, the orifice pressure drops become functions of p_L , and q_{in} becomes a function of the input displacement x_1 and the load pressure drop p_L :

$$q_{in} = f(x_1, p_L) \quad (5-5-13)$$

The incremental approach of Chapter 3 is useful. The total differential of input flow rate is

$$dq_{in} = \frac{\partial q_{in}}{\partial x_1} dx_1 + \frac{\partial q_{in}}{\partial p_L} dp_L \quad (5-5-14)$$

If the partial derivatives are considered to be constant over the operating range of x_1 and p_L about a selected operating point, Eq. 5-5-14 is a linear relationship for increments. Since normal operating points are $x_1 = 0$ and $p_L = 0$, the increments are actually total values of the variables. As long as only small values of x_1 and p_L are used, Eq. 5-5-14 can be written as

$$q_{in} \approx K_1 x_1 + K_2 p_L \quad (5-5-15)$$

In this equation

$$K_1 = \left. \frac{\partial q_{in}}{\partial x_1} \right|_{p_L = 0} \quad (5-5-16)$$

$$K_2 = \left. \frac{\partial q_{in}}{\partial p_L} \right|_{x_1 = \text{constant} \approx 0} \quad (5-5-17)$$

The constant K_2 is numerically negative since input flow rate decreases for an increase in load pressure at a constant valve opening.

The effect of a leakage flow rate can be found as follows. In general leakage depends on load pressure and experiments indicate that a leakage flow rate term, q_l , directly proportional to p_L is suitable. This relationship can be expressed as

$$q_l = K_l p_L \quad (5-5-18)$$

The flow rate established by movement of the actuator is

$$q_a = A_p \frac{dx_2}{dt} \quad (5-5-19)$$

Finally the input flow rate must equal the sum of q_a and q_l ,

$$q_{in} = K_1 x_1 + K_2 p_L = K_l p_L + A_p \frac{dx_2}{dt} \quad (5-5-20)$$

Before Eq. 5-5-20 can be altered to relate only x_1 and x_2 , a relationship between p_L and the actuator displacement, x_2 , must be written. This depends on the type of load on the actuator. For a viscous friction and mass load

$$A_p p_L = F_L = M \frac{d^2 x_2}{dt^2} + f_v \frac{dx_2}{dt} \quad (5-5-21)$$

Eqs. 5-5-20 and 5-5-21 can be combined and simplified as follows,

$$K_1 x_1 = (K_l - K_2) \frac{M}{A_p} \frac{d^2 x_2}{dt^2} + \left[A_p + (K_l - K_2) \frac{f_v}{A_p} \right] \frac{dx_2}{dt} \quad (5-5-22)$$

A transfer function can now be written as

$$\frac{X_2(s)}{X_1(s)} = \frac{K_1 A_p / M (K_l - K_2)}{s \{ s + [A_p^2 / M (K_l - K_2) + f_v / M] \}} \quad (5-5-23)$$

The effect of compressibility is still not included in the transfer function of Eq. 5-5-23. However the approximate effects of load pressure and leakage on the input flow rate are added to the simplified analysis of Ex. 5-5-1.

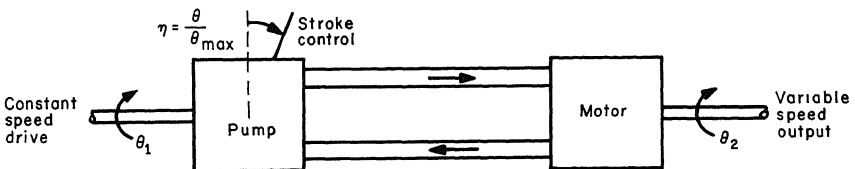


Fig. 5-5-3.

Example 5-5-4. The hydraulic pump controlled motor of Fig. 5-5-3 is a slightly different hydraulic actuator system that gives rotary motion as an output. This system provides a final example for obtaining a hydraulic system transfer function. The pump used in this combination is a variable stroke piston pump that uses a tiltable plate to control the stroke of the pistons.² The tilt plate position is used as the input; however, it is convenient to use the fraction of maximum tilt plate displacement as the actual input variable. The hydraulic motor that is used has a fixed displacement; therefore variation of fluid flow through the motor causes a direct variation in motor speed.

A reasonably detailed analysis of the hydraulic pump controlled motor can be made by finding the flow rate through the pump and equating this to the sum of actual motor flow rate, leakage flow rate, and compression flow rate. The flow rate through the pump, q_p , must be directly proportional to the piston stroke. Since pump piston stroke is proportional to tilt plate position,

$$q_p = K_p \eta \quad (5-5-24)$$

where η is the input expressed as a fraction of maximum tilt plate angle. The actual volume of fluid flowing through the motor and causing rotation depends on the fixed piston displacement, the number of pistons, and the number of strokes. Since there is generally one stroke of each piston per shaft revolution, the volume of flow through the motor is proportional to motor shaft angle and the motor flow rate, q_m , is proportional to motor velocity,

$$q_m = K_m \frac{d\theta_2}{dt} \quad (5-5-25)$$

The leakage flow rate for the system which is being considered is taken as proportional to the drop in pressure, p_L , across the motor,

$$q_l = K_l p_L \quad (5-5-26)$$

The final flow rate to be considered is that due to compression effects. From the definition of the bulk modulus of fluids in Art. 5-2 it is possible to establish a compression flow rate. The approach used here follows that of Gibson and Tuteur.³ For a blocked motor and equal pressures in two equal volume lines, the effect of pumping of fluid, Q , from one line to another raises the pressure in one line by Δp and lowers the pressure in other line by the same amount. The relationship between Q , Δp , and B is

$$Q = \frac{V}{2} \frac{\Delta p}{B} \quad (5-5-27)$$

where V is the total volume of fluid in the system. After the fluid has been pumped from one line to another, the difference in pressure between lines is $2\Delta p = p_L$. Therefore

$$Q = \frac{V}{4B} p_L \quad (5-5-28)$$

² John E. Gibson and Franz B. Tuteur, *Control System Components*, McGraw-Hill Book Co., New York, Chapter 10.

³ *Ibid.*

The compression volume flow rate, q_c , then becomes

$$q_c = \frac{dQ}{dt} = \frac{V}{4B} \frac{dp_L}{dt} = K_c \frac{dp_L}{dt} \quad (5-5-29)$$

There is no compression flow rate unless the load pressure is changing. The pump flow rate can now be equated to the sum of the three flow rate components, q_m , q_i , and q_c as follows.

$$K_p \eta = K_m \frac{d\theta_2}{dt} + K_i p_L + K_c \frac{dp_L}{dt} \quad (5-5-30)$$

In order to continue the development of an over-all transfer function, a load pressure-to-developed torque equation plus an output torque to output angular displacement equation must be obtained. The developed torque is proportional to the drop in pressure across the motor.

$$T_D = K_m p_L \quad (5-5-31)$$

The constant K_m here is the same as that in Eq. 5-5-25. This fact is obtained by equating hydraulic and mechanical forms for power.

$$\text{Hydraulic power} = q_m p_L = K_m \frac{d\theta_2}{dt} p_L \quad (5-5-32)$$

$$\text{Mechanical power} = T_D \frac{d\theta_2}{dt} = K_{\text{torque}} p_L \frac{d\theta_2}{dt} \quad (5-5-33)$$

$$K_{\text{torque}} = K_m \quad (5-5-34)$$

For a general load torque equation which accounts for moment of inertia, viscous friction, and a spring.

$$T_L = J \frac{d^2 \theta_2}{dt^2} + D \frac{d\theta_2}{dt} + K_s \theta_2 = T_D \quad (5-5-35)$$

By combining Eqs. 5-5-31 and 5-5-35

$$p_L = \frac{1}{K_m} \left[J \frac{d^2 \theta_2}{dt^2} + D \frac{d\theta_2}{dt} + K_s \theta_2 \right] \quad (5-5-36)$$

Finally Eq. 5-5-36 can be substituted into Eq. 5-5-30 to obtain the desired differential equation,

$$K_m K_p \eta = (K_c J) \frac{d^3 \theta_2}{dt^3} + (K_i J + K_c D) \frac{d^2 \theta_2}{dt^2} + (K_m^2 + K_i D + K_c K_s) \frac{d\theta_2}{dt} + (K_i K_s) \theta_2 \quad (5-5-37)$$

The system transfer function is

$$G(s) = \frac{\Theta_2(s)}{\eta(s)} = \frac{K_m K_p / K_c J}{s^3 + \left(\frac{K_i}{K_c} + \frac{D}{J} \right) s^2 + \left(\frac{K_m^2}{K_c J} + \frac{K_i D}{K_c J} + \frac{K_s}{J} \right) s + \frac{K_i K_s}{K_c J}} \quad (5-5-38)$$

The complexity of this transfer function prevents much general discussion of the

system characteristics. However, application of the final-value theorem for a unit-step input gives a final output angle,

$$\theta_{2ss} = \frac{K_m K_p}{K_i K_s} \quad (5-5-39)$$

For this condition the pump flow rate is K_p , and the output torque is $\theta_{2ss} K_s$ which requires a load pressure $\theta_{2ss} K_s / K_m$. Since the motor is blocked and the load pressure is constant, all of the motor flow must be leakage flow. This is shown by rearranging Eq. 5-5-39 as follows and by using the expressions for pump flow and load pressure,

$$q_{lss} = K_l p_{Lss} = K_l \left(\frac{\theta_{2ss} K_s}{K_m} \right) = K_p \quad (5-5-40)$$

This equation may not be valid since it may not be possible to obtain the load pressure p_{Lss} with practical components. The supply lines might rupture, the pump might stall, etc. Thus equations do not tell everything. They must be used with care and judgment. In fact, the load pressure variation as well as the output position should be checked in each problem to indicate excessive pressure. Furthermore, if the spring load is absent, the system becomes an integrating device. A fixed input stroke causes a fixed output rate of change.

5-6. General Pneumatic System Relations. The basic parameters and variables used in air systems are the same as those used in regular hydraulic systems. However certain additional factors must be considered. First among these is the role played by compressibility. In many low pressure, small volume hydraulic systems compressibility can be neglected; however compression of air is an important factor in pneumatic systems. The type of expansion or compression in air systems is a second important factor which is not considered in hydraulic systems. In most cases the air process is intermediate between an adiabatic (no heat exchange) and an isothermal process. A third new factor commonly encountered in air systems is supersonic flow through orifices and restrictions. This condition occurs when the pressure drop across an orifice or restriction exceeds 47 per cent of the upstream pressure. The flow rate is independent of the downstream pressure under such conditions.

Three basic flow equations are needed for air systems—one for subsonic flow through orifices, one for supersonic flow through orifices, and one for smooth flow through tubes and capillaries. The first of the required equations is

$$w = \frac{C_d A_2}{v_2} \sqrt{\frac{2gn}{n-1}} p_1 v_1 \left[1 - \left(\frac{p_2}{p_1} \right)^{(n-1)/n} \right] \quad (5-6-1)$$

which assumes that the input velocity can be neglected compared to the velocity through the restriction. In Eq. 5-6-1 the symbols have the following meanings:

w = weight flow rate — lbs/sec
 C_d = discharge coefficient

A_2 = area of orifice of restriction — in².

v_1 & v_2 = upstream and downstream specific volumes respectively
— in.³/lb

p_1 & p_2 = upstream and downstream pressures, respectively — lbs/in²
 n = polytropic process exponent (1.4 for adiabatic, 1.0 for isothermal)

g = acceleration of gravity — in./sec²

By using a binomial expansion of the factor $[1 - (p_2/p_1)^{(n-1)/n}]$, the following simplification is possible.

$$1 - \left(\frac{p_2}{p_1}\right)^{(n-1)/n} = 1 - \left(1 - \frac{p_1 - p_2}{p_1}\right)^{(n-1)/n} \quad (5-6-2)$$

$$= 1 - \left[1 - \left(\frac{n-1}{n}\right)\left(\frac{p_1 - p_2}{p_1}\right) + \dots\right] \quad (5-6-3)$$

$$\approx \left(\frac{n-1}{n}\right)\left(\frac{p_1 - p_2}{p_1}\right); (p_1 - p_2) \ll p_1 \quad (5-6-4)$$

Flow eq. 5-6-1 can therefore be approximated as

$$w = \frac{C_a A_2}{v_2} \sqrt{2g v_1 (p_1 - p_2)} \quad (5-6-5)$$

as long as the orifice pressure drop is a small percentage of the input pressure. Eq. 5-6-1 gives a maximum value for adiabatic flow at

$$p_2 = 0.53 p_1 \quad (5-6-6)$$

This leads to the second pneumatic flow equation. For $p_2 < 0.53 p_1$ supersonic flow exists through orifices and

$$w = \frac{0.53 C_a A_2 p_1}{\sqrt{T_1}} \quad (5-6-7)$$

where T_1 is the temperature of input air in degrees Rankine. The weight flow rate is independent of the drop in pressure under these conditions. The third flow equation is for smooth flow through tubes. This equation is

$$w = \frac{\pi D^4 (p_1^2 - p_2^2)}{256 \mu l v_1 p_1} \quad (5-6-8)$$

in which D is tube diameter, μ is absolute viscosity of air, and l is tube length. For small drops in pressure

$$w = \frac{\pi D^4 (p_1 - p_2)}{128 \mu l v_1} \left(1 + \frac{p_2}{p_1} \right) \left(\frac{1}{2} \right) \quad (5-6-9)$$

$$\approx \frac{\pi D^4 (p_1 - p_2)}{128 \mu l v_1} \quad (5-6-10)$$

Eq. 5-6-10 is a linear equation but is subject to considerable error if p_2 is much smaller than p_1 . All of the original flow relations above are highly nonlinear. Thus a linear approximation approach must be used to simplify the analyses that follow.

5-7. Pneumatic Resistance and Capacitance. The concepts of pneumatic resistance and capacitance can be used in a manner somewhat similar to hydraulic resistance and capacitance. Incremental pneumatic resistance is defined as

$$R_p = \frac{d(p_1 - p_2)}{dw} \quad (5-7-1)$$

It is therefore the slope of the curve relating pressure drop to weight flow rate. For flow orifices when the drop in pressure is small, Eq. 5-6-5 can be used in an approximate manner (v_2 and v_1 considered constant at operating values) to give

$$\frac{1}{R_p} = \frac{C_d A_2}{v_2} \sqrt{2g v_1} \frac{1}{2} \frac{1}{\sqrt{p_1 - p_2}} \quad (5-7-2)$$

$$R_p = \frac{2(p_1 - p_2)}{w} \quad (5-7-3)$$

This resistance can be evaluated at the operating values of w and $(p_1 - p_2)$ and treated as a constant for small variations from the operating point. If the normal operating point is zero flow and zero pressure drop, R_p is zero and no useful result is obtained. In such a case an average R_p determined by the maximum excursion of $p_1 - p_2$ and w should be used. Fig. 5-7-1 illustrates such a choice.

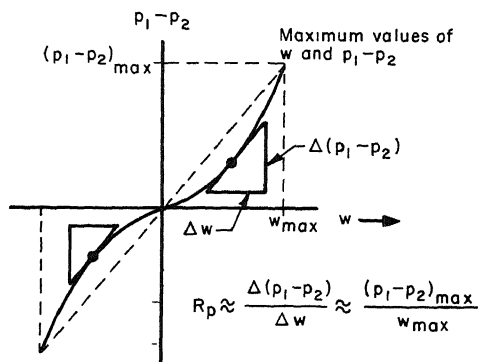


Fig. 5-7-1.

Example 5-7-1. The air resistance for a 0.1 in. diameter orifice with a discharge coefficient of 0.62 is to be found assuming a small pressure drop and

$$\left. \begin{aligned} p_1 &= 15 \text{ psig} \\ p_2 &= 10 \text{ psig} \\ T_1 &= 68^\circ \text{ F.} \end{aligned} \right\} \text{operating values}$$

$$\rho = \frac{1}{v} = 0.0753 \text{ lbs/ft}^3 \text{ at } 68^\circ \text{ F. and } 14.7 \text{ psia}$$

The operating weight flow rate determination requires v_1 and v_2 . For an approximately isothermal process

$$p_1 v_1 = \text{constant} = p_2 v_2 = (pv)_{14.7 \text{ psia}} \quad (5-7-4)$$

Therefore

$$v_1 = \left(\frac{14.7}{14.7 + 15} \right) \left(\frac{1}{0.0753} \right) = 6.48 \text{ ft}^3/\text{lb} \quad (5-7-5)$$

$$= (6.48)(1728) = 1.12 \times 10^4 \text{ in.}^3/\text{lb} \quad (5-7-6)$$

$$v_2 = \left(\frac{17.4}{14.7 + 10} \right) \left(\frac{728}{0.0753} \right) = 1.365 \times 10^4 \text{ in.}^3/\text{lb} \quad (5-7-7)$$

At the operating point Eq. 5-6-5 is

$$w_o = \frac{(0.62)(\pi \times 10^{-2}) \sqrt{(2)(32.2 \times 12)(1.12 \times 10^4)(15 - 10)}}{(4)(1.365 \times 10^4)} \quad (5-7-8)$$

$$w_o = 2.35 \times 10^{-3} \text{ lb/sec} \quad (5-7-9)$$

The pneumatic resistance is obtained from Eq. 5-7-3 as

$$R_p = \frac{2(15 - 10)}{2.35 \times 10^{-3}} = 4.26 \times 10^3 \text{ sec/in.}^2 \quad (5-7-10)$$

Example 5-7-2. In this example a suitable pneumatic resistance of the orifice of Ex. 5-7-1 is to be found when the operating weight flow rate is zero. However, pressure drop variations up to 5 psia are expected. The pressure p_1 is assumed to be constant at 15 psig. Eq. 5-7-2 reduces to zero in this case. The actual weight flow

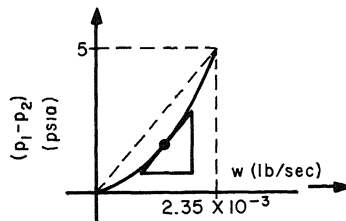


Fig. 5-7-2.

rate curve is shown in Fig. 5-7-2. The pneumatic resistance can be taken as shown in Fig. 5-7-1.

$$R_p \approx \frac{(p_1 - p_2)_{\max}}{w_{\max}} = \frac{5}{2.35 \times 10^{-3}} = 2.13 \times 10^3 \text{ sec/in.}^2 \quad (5-7-11)$$

For cases in which air flows through capillary tubes in a system, Eq. 5-6-8 is generally applicable. However the form of this equation makes it difficult to find R_p . Eq. 5-6-10 provides a suitable approximation for determining R_p in such cases.

$$R_p \approx \frac{128\mu v_1}{\pi D^4} \quad (5-7-12)$$

Example 5-7-3. Eq. 5-7-12 can be used to find the pneumatic resistance of a 0.05 in. diameter capillary tube 4 in. long at 68° F as follows. For air $\mu = 2.5 \times 10^{-9}$ $(T/273)^{3/4}$ lbs-sec/in.² (T in degrees Kelvin). An input pressure of 20 psig is assumed. The following quantities must be determined first.

$$\mu = (2.5 \times 10^{-9}) \left(\frac{293}{273} \right)^{3/4} = 2.64 \times 10^{-9} \text{ lb-sec/in.}^2 \quad (5-7-13)$$

$$v_1 = \left(\frac{14.7}{34.7} \right) \left(\frac{1728}{0.0753} \right) = 9.71 \times 10^3 \text{ in.}^3/\text{lb} \quad (5-7-14)$$

R_p can now be found as

$$R_p \approx \frac{(128)(2.64 \times 10^{-9})(4)(9.71 \times 10^3)}{\pi(5 \times 10^{-2})^4} \quad (5-7-15)$$

$$\approx 662 \text{ sec/in.}^2 \quad (5-7-16)$$

The difficulty encountered in evaluating R_p at $p_1 - p_2 = 0$ in Ex. 5-7-2 is not present when using Eq. 5-7-12.

The concept of pneumatic capacitance is similar to that of hydraulic capacitance in Art. 5-4. However, for pressurized vessels pneumatic capacitance depends on the type of expansion process involved. Pneumatic capacitance is defined by

$$C_p = \frac{dW_{\text{total}}}{dp} \quad (5-7-17)$$

where W_{total} is the total weight of air in a pressurized volume. Since total weight is total volume times specific weight, C_p for essentially constant volumes can be written as

$$C_p = (\text{volume}) \frac{d\varrho}{dp} \quad (5-7-18)$$

in which ϱ is specific weight. From thermodynamics

$$pv^n = p/\varrho^n = \text{constant} \quad (5-7-19)$$

$$pv = p/\varrho = R_{\text{gas}} T \quad (5-7-20)$$

The term R_{gas} is the universal gas constant. C_p can now be found by differentiating Eq. 5-7-19 and substituting from Eq. 5-7-20.

$$C_p = \frac{\text{Volume}}{nR_{\text{gas}} T} \quad (5-7-21)$$

Example 5-7-4. To illustrate pneumatic resistance and capacitance, the capillary-volume system of Fig. 5-7-3 is analyzed to obtain the transfer function which relates increments in supply pressure $p_{s\Delta}$ to increments in pressure $p_{1\Delta}$ in the

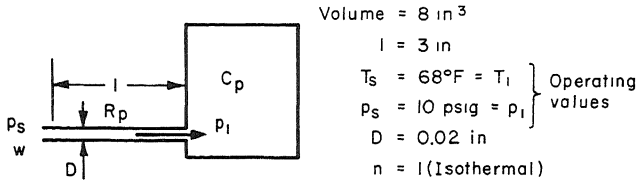


Fig. 5-7-3.

vessel. The pneumatic resistance, R_p , of the capillary is determined as follows according to the procedure of Ex. 5-7-3 from which example $\mu = 2.64 \times 10^{-9}$ lb-sec/in. From Eq. 5-7-5, $v_1 = 1.365 \times 10^4$ in.³/lb

$$R_p \approx \frac{(128)(2.64 \times 10^{-9})(3)(1.365 \times 10^4)}{\pi(2)^4(10^{-8})} \quad (5-7-22)$$

$$\approx 276 \times 10^3 \text{ sec/in.}^2 \quad (5-7-23)$$

$$C_p = \frac{8}{(1)(53.3)(526)(12)} \frac{\text{in.}^3}{(\text{ft}/^\circ\text{Rankine})(^\circ\text{Rankine})(\text{in.}/\text{ft})} \quad (5-7-24)$$

$$= 2.37 \times 10^{-4} \text{ in.}^2 \quad (5-7-25)$$

The equation relating the increment in weight flow rate w_Δ through the capillary to the difference in increments of pressure $p_{s\Delta} - p_{1\Delta}$ is

$$p_{s\Delta} - p_{1\Delta} = R_p w_\Delta \quad (5-7-26)$$

The relation between the time variation of increments of pressure, $p_{1\Delta}$, in the vessel and the weight flow rate increment is

$$C_p \frac{dp_{1\Delta}}{dt} = w_\Delta \quad (5-7-27)$$

In reconciling Eqs. 5-7-17 and 5-7-27 it should be noted that w is a rate of flow and not a total weight. Eqs. 5-7-26 and 5-7-27 can be combined to give

$$R_p C_p \frac{dp_{1\Delta}}{dt} + p_{1\Delta} = p_{s\Delta} \quad (5-7-28)$$

$$(276 \times 10^3)(2.37 \times 10^{-4}) \frac{dp_{1\Delta}}{dt} + p_{1\Delta} = p_{s\Delta} \quad (5-7-29)$$

$$\frac{dp_{1\Delta}}{dt} + \frac{p_{1\Delta}}{65.4} = \frac{p_{s\Delta}}{65.4} \quad (5-7-30)$$

The desired transfer function can be written by inspection as

$$G(s) = \frac{P_{1\Delta}(s)}{P_{s\Delta}(s)} = \frac{1.53 \times 10^{-2}}{s + 0.0153} \quad (5-7-31)$$

The system which is being considered is characterized by a single pole at $s = -0.0153$. Thus the impulse response has a single time constant of 65.4 sec. This long time constant is typical of some pneumatic components.

Example 5-7-5. The simple capillary-bellows system of Fig. 5-7-4 is another example of pneumatic components. A transfer function relation between the posi-

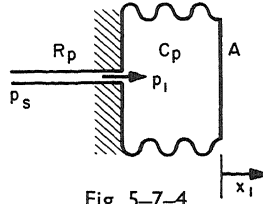


Fig. 5-7-4.

tion x_1 and increments in p_s is needed. The position x_1 is chosen as zero when p_s is at normal operating pressure. The bellows acts as a spring so that

$$\text{Force} = K_s x_1 = A p_{1\Delta} \quad (5-7-32)$$

The term A is the effective bellows area upon which pressure acts to elongate the bellows. The volume of expansion of the bellows is assumed small so that C_p is essentially constant.

From Ex. 5-7-4

$$\frac{P_{1\Delta}(s)}{P_{s\Delta}(s)} = \frac{1/R_p C_p}{s + 1/R_p C_p} \quad (5-7-33)$$

Eq. 5-7-32 can be transformed to give

$$\frac{X_1(s)}{P_{1\Delta}(s)} = \frac{A}{K_s} \quad (5-7-34)$$

Eqs. 5-7-33 and 5-7-34 are then combined to give the desired transfer function.

$$\frac{X_1(s)}{P_{s\Delta}(s)} = \frac{A/K_s R_p C_p}{s + 1/R_p C_p} \quad (5-7-35)$$

5-8. The General Incremental Analysis of Pneumatic Systems.

Many pneumatic problems require more than just R_p and C_p . The general incremental approach of Chapter 3 must be applied. Since this method is best illustrated by example, its development is made in the example that follows.

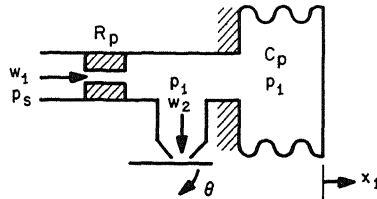


Fig. 5-8-1.

Example 5-8-1. Fig. 5-8-1 illustrates a nozzle, baffle plate, bellows, restriction, and constant supply pressure p_s . A transfer function is to be found for $X_1(s)/\Theta(s)$. The bellows and nozzle are assumed to have a constant capacity C_p . The restriction has a resistance R_p . Other assumptions are that supersonic flow occurs through

the nozzle and that only small variations from normal operating values exist. The variables x and θ are chosen as zero at normal operating values of w_2 and p_1 . Since supersonic flow through the nozzle is assumed to exist

$$w_2 = f(\theta, p_1) \quad (5-8-1)$$

$$dw_2 = \frac{\partial w_2}{\partial \theta} d\theta + \frac{\partial w_2}{\partial p_1} dp_1 \quad (5-8-2)$$

The partial derivatives are evaluated at the operating values of w_2 , θ , and p_1 and then considered to be constants for the small excursions that occur in operation.

$$K_1 = \left. \frac{\partial w_2}{\partial \theta} \right|_{p_1 - \text{constant}} \quad (5-8-3)$$

$$K_2 = \left. \frac{\partial w_2}{\partial p_1} \right|_{\theta=0} \quad (5-8-4)$$

Eq. 5-8-2 can now be written as

$$w_{2\Delta} = K_1\theta + K_2p_{1\Delta} \quad (5-8-5)$$

The choice of the position for $\theta = 0$ means that increments from the normal position are actually equal to the total value of θ . From Ex. 5-7-4 and 5-7-5 and the conservation of weight of air

$$p_{s\Delta} - p_{1\Delta} = R_p w_{1\Delta} \quad (5-8-6)$$

$$C_p \frac{dp_{1\Delta}}{dt} = w_{1\Delta} - w_{2\Delta} \quad (5-8-7)$$

$$K_s x_1 = A p_{1\Delta} \quad (5-8-8)$$

Eqs. 5-8-5, 5-8-6, and 5-8-7 can be combined to eliminate $w_{1\Delta}$ and $w_{2\Delta}$.

$$C_p \frac{dp_{1\Delta}}{dt} = \frac{1}{R_p} (p_{s\Delta} - p_{1\Delta}) - (K_1\theta + K_2p_{1\Delta}) \quad (5-8-9)$$

$$\frac{dp_{1\Delta}}{dt} + \left(\frac{1}{R_p C_p} + \frac{K_2}{C_p} \right) p_{1\Delta} = \frac{p_{s\Delta}}{R_p C_p} - \frac{K_1}{C_p} \theta \quad (5-8-10)$$

The pressure $p_{1\Delta}$ is now eliminated to give

$$\frac{dx_1}{dt} + \left(\frac{1}{R_p C_p} + \frac{K_2}{C_p} \right) x_1 = \frac{A}{K_s R_p C_p} p_{s\Delta} - \frac{AK_1}{K_s C_p} \theta \quad (5-8-11)$$

For the constant supply pressure assumed originally, $p_{s\Delta} = 0$ and the transfer function can be written as

$$\frac{X_1(s)}{\Theta(s)} = \frac{-AK_1/K_s C_p}{\left[s + \left(\frac{1}{R_p C_p} + \frac{K_2}{C_p} \right) \right]} \quad (5-8-12)$$

The presence of the minus sign in Eq. 5-8-12 indicates that positive values of θ lead to a decrease in x_1 . This is caused by an increase in w_2 and w_1 which in turn causes a greater drop in pressure across R_p . The effect of an increment in p_s can be seen by studying Eq. 5-8-11. Positive increments in p_s lead to positive increments in x_1 through a single time constant relationship.

5-9. Conclusion. This chapter reviews hydraulic and pneumatic equations and introduces various methods of analyzing such systems. Because of the nonlinearities involved, the use of the incremental method predominates. Even its accuracy is questionable in many cases. The concepts of hydraulic and pneumatic resistance and capacitance are introduced to aid in formulating system equations. In reality they are analogous to electrical resistance and capacitance. This analogy was not stressed since a later chapter deals with analog methods and since it would introduce more material, possibly detracting from the immediate analysis problems. The s -plane pole-zero plots for transfer functions have not been drawn for systems in this chapter; however, they should not be forgotten as they provide a common point for comparing the dynamic characteristics of hydraulic and pneumatic systems with mechanical and electrical systems. Later chapters tend to overlook the actual physical origin of the s -plane pole-zero plots and study the characteristics of forms of transfer functions. A knowledge of the transfer functions of various systems and components is a very useful aid in visualizing what happens physically.

PROBLEMS

5-1. Calculate the laminar hydraulic resistance of a 1 in. diameter pipe 10 ft long for water flow at 68°F.

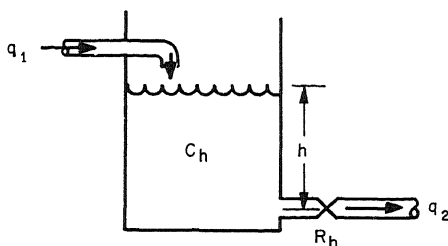


Fig. 5-P-2.

5-2. For the system of Fig. 5-P-2 determine the response in h to a step change in q_1 of $1/80$ ft³/sec from its operating value. The normal steady operating head is 4 ft and the normal operating value of $q_2 = q_1 = 1/8$ ft³/sec. The vessel is cylindrical and 6 ft² in cross-sectional area. Assume R_h is the resistance of an orifice.

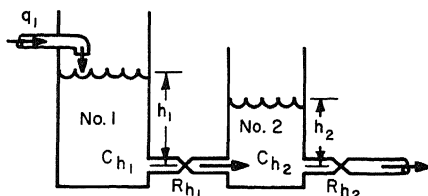


Fig. 5-P-3.

5-3. Determine the s -plane pole-zero pattern for the system of Fig. 5-P-3 if q_1 is considered the input and h_2 is the output. Tank no. 1 is cylindrical and 2 ft in diameter while tank no. 2 is rectangular in cross section with sides 1 ft by 2 ft. Restriction no. 1 is 0.7 in.^2 in area and has a discharge coefficient of 0.65. Restriction no. 2 is 0.8 in.^2 in area and has a discharge coefficient of 0.65. The operating head in tank no. 2 is 2 ft. Other operating values are such as to maintain a steady value of head in tank no. 1.

5-4. Discuss the effect of the ratio l_1/l_2 on the response of the idealized valve operated piston of Ex. 5-5-2.

5-5. Determine the time varying pressure drop across the idealized piston of Ex. 5-5-2 if a load of $f_v = 200 \text{ lb-sec/ft}$ and $K_s = 100 \text{ lb/in.}$ is moved horizontally due to a 0.2 in. step input x_3 . The piston effective area is 2 in.^2 . A 0.1 in. input movement gives a 2 in. output with 63.2% of the change accomplished in 0.4 sec.

5-6. Reanalyze the valve operated piston system of Ex. 5-5-3 by including a compression flow rate, $q_c = K_c dp_L/dt$.

5-7. Obtain a transfer function which relates actuator piston output position to pilot valve input position assuming that the pilot valve is an ideal flow control valve ($q_{in} = Kx_1$). Include compressibility and leakage effects if they don't drop out automatically.

5-8. Repeat Prob. 5-7 for an ideal pressure control valve ($p_L = Kx_{in}$). What happens to compression and leakage effects?

5-9. A pilot valve actuated piston as shown in Fig. 5-5-1 has the following characteristics:

Actuator piston area = 0.96 in.^2

„ „ stroke = 4 in.

„ „ load weight = 140 lbs

„ „ load damping = 10 lbs-sec/in.

„ „ spring load = 2000 lbs/in.

Leakage flow versus load pressure with a blocked actuator is shown in Fig. 5-P-9. The supply pressure is 3,000 psi. The rate of change of input flow rate vs. pilot

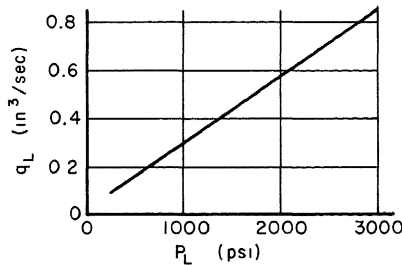


Fig. 5-P-9.

valve position at constant load pressure is $4000 \text{ (in.}^3/\text{sec)/in.}$ The rate of change of input flow rate vs. load pressure for constant input position is $-4 \times 10^{-4} \text{ in.}^5/(\text{lb-sec})$. Determine the transfer function Eq. 5-5-21 for this combination neglecting compressibility.

5-10. The valve-piston system of Prob. 5-9 has an over-all tubing, manifold, and piston volume of 16 in.^3 . Assume a bulk modulus for the hydraulic fluid of 2.25

$\times 10^5$ psi. A $10^0\%$ expansion in tubing is to be allowed for by a change in the bulk modulus. Using in addition, information in Prob. 5-6 and 5-9 and the K_c equation 5-5-9, find the transfer function for the valve-piston system.

5-11. Find the steady-state response of the valve-piston system of Prob. 5-10 to an input

$$x_1 = 0.001 \sin 300t \text{ in.}$$

5-12. Determine an expression which relates the leakage coefficient of the pump controlled hydraulic motor of Ex. 5-5-4 to steady-state change in output rpm divided by the torque increment required to cause the speed change. Note that an increase in torque causes a decrease in speed due to the diversion of fluid thru leakage paths.

5-13. A piston-type hydraulic pump with a tilt plate has the following ratings:

Displacement = $1 \text{ in.}^3/\text{rev max}$

Speed = 4,000 rpm

Rated pressure = 3,000 psi

A motor identical to the pump except for a fixed displacement of $1 \text{ in.}^3/\text{rev}$ is driven by the pump. The total leakage is proportional to load pressure and is $15^0\%$ of the actual motor flow at rated pressure of 3,000 psi. The pump and motor are connected by $1/2$ in. diameter tubing and they are 25 ft apart. Allow a $15^0\%$ increase in the compressibility coefficient to account for tubing expansion. Assume $B = 2.5 \times 10^5$ psi. The motor is loaded by a weight of 25 lbs which has a radius of gyration of 2 in. Additionally, viscous friction $D = 1.75 \text{ in.-lb-sec}$ exists. Neglect line pressure drops. Determine the transfer function which relates motor speed to percentage tilt plate motion.

5-14. Find the transfer function which relates the load pressure, p_{Ls} to the percentage input tilt plate motion for the pump-controlled motor of Prob. 5-13. Find the steady-state load pressure for full $100^0\%$ tilt plate movement.

5-15. Calculate the air resistance of a 0.08 in. diameter orifice with a discharge coefficient of 0.65 for

$$\left. \begin{array}{l} p_1 = 10 \text{ psig} \\ p_2 = 8 \text{ psig} \\ T_2 = T_1 = 100^\circ \text{ F} \end{array} \right\} \text{ operating values}$$

5-16. Plot a curve of the pneumatic resistance of a capillary tube 6 in. long and 10 mils in diameter for temperatures from 0°F to 200°F . Assume no change in diameter due to temperature. Assume an isothermal process. Use semilog graph paper with temperature on the log scale.

5-17. Plot the pneumatic capacitance for a 2 in.^3 volume vs. temperature from 0° to 200°F for an adiabatic process. Repeat for an isothermal process. Use semilog graph paper with temperature on the log scale.

5-18. Plot the isothermal pneumatic time constant, $R_p C_p$, vs. temperature for $0^\circ < T < 200^\circ\text{F}$ for the R_p of Prob. 5-15 and C_p of Prob. 5-17. Use semilog graph paper with temperature on the log scale.

5-19. The system of Fig. 5-8-1 has the following characteristics.

$$\begin{array}{l} R_p = 600 \text{ sec/in.}^2 \\ C_p = 2 \times 10^{-3} \text{ in.}^2 \end{array}$$

At a normal operating value of $p_1 = 10$ psig a change of 0.2 degrees in θ causes an increase of 3×10^{-3} lb/sec in w_2 . At $\theta = 0$ an increase in p_1 of 1 psi causes an increase of 1×10^{-3} lb/sec in w_2 . The bellows has an effective cross sectional area of 20 in.² and an effective spring constant of 500 lbs/in. Find the approximate transfer function $X_1(s)/\Theta(s)$.

6

ANALYSIS OF SIMPLE CONTROL SYSTEMS

6-1. Introduction. This chapter applies the mathematical procedures developed in Chapters 2, 3, 4, and 5 to simple control systems. The first part of this analysis is concerned with the transient response, the second determines the steady-state response when a sinusoidal input is applied to the system.

If the input to a control system is changed to a new function, the system goes through a transient period and if the system is so designed, settles into a steady-state response. Nevertheless, the variations during the transient period may be quite pronounced and therefore need to be investigated. The transient analysis of control systems determines the complete response of the system to some designated input in that this solution contains both the particular and the complementary component of the solutions.

Steady-state sinusoidal analysis of control systems assumes that a sinusoidal function is applied to the input and is maintained long enough for the transient (or complementary) component of the solution to decay leaving only the steady-state (or particular) component of the solution. This analysis is analogous to a-c circuit theory as discussed in Chapter 3.

6-2. Types of Input Signals. Because of the wide variety of inputs that control systems may be expected to follow, it may be impossible to analyze a system under actual input conditions. However, the actual input can usually be interpreted so that test inputs can be applied and, if the system response is satisfactory, the system response to the actual input is also satisfactory.

The problem of test inputs is not unique with control systems and a review of a more familiar situation may be reassuring. For example, an audio amplifier can be connected to the output of an AM tuner and the input signal to the amplifier originates from whatever program the radio station is broadcasting at the moment. This amplifier was not tested with this exact signal. However, in all likelihood, a sinusoidal test input of varying frequency was applied to the amplifier. If it passed a certain band of frequencies, it was assumed to be satisfactory.

The control system designer can examine the general nature of the input to which the system is to respond. He can then choose a test input more severe than the expected input. The impulse family of functions discussed in Chapter 2 has proved very useful as test inputs. The unit-step function represents a sudden change in the input and places rather severe demands on the system. A unit-ramp function represents the type of input followed by a sweep-radar and similar systems. In other situations other members of this family may be useful. If the response of the system to these test inputs is satisfactory, the system will certainly handle the expected input.

The other type of input function in common use is the sinusoid. As mentioned previously, the calculation of the steady-state response to a sinusoid of varying frequency is analogous to calculation of the frequency response curve for an amplifier.

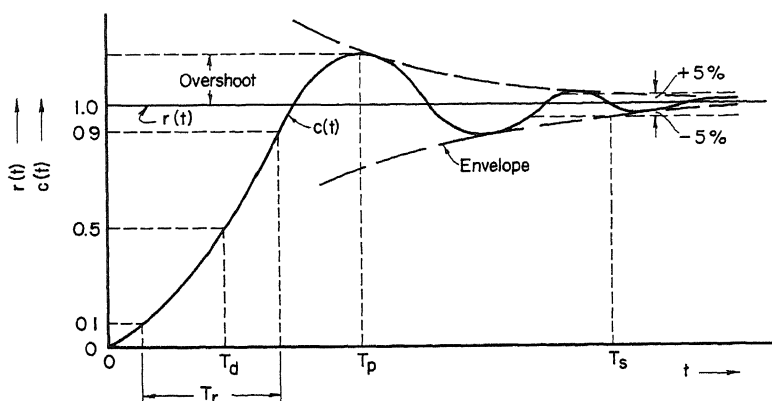


Fig. 6-3-1.

6-3. Nomenclature for Step Function Response. Certain nomenclature exists concerning a system's response to a unit-step function. A typical response is shown in Fig. 6-3-1 as the $c(t)$ curve. Much of the information contained in this curve can be summarized by 5 terms:

1. The overshoot is the maximum value by which the output exceeds the input and is usually given as a percentage of the final value.
2. The time required for $c(t)$ to reach its maximum value is denoted by T_p .
3. The time required for the output $c(t)$ to rise from 10% to 90% of the final value is called the rise time and is denoted by T_r .
4. The time required for the output $c(t)$ to reach 50% of the final value is called delay time and is denoted by T_d .
5. The time required for the output $c(t)$ to settle and remain within $\pm 5\%$ of its final value is called settling time and is denoted by T_s . In order to have a continuously variable settling time for design purposes, it is more convenient to consider the envelope instead of the actual curve.

In some cases different percentage values are used for rise time and for settling time. For example, $\pm 2\%$ of the final value is sometimes used for settling time. However, even though the percent may be different, the terms mean the same thing.

6-4. Nomenclature for Sinusoidal Steady State Response.

Certain nomenclature exists concerning a system's steady-state response to an input that is a sinusoid of varying frequency. The magnitude M of a typical output response $|C(j\omega)|$ when the magnitude of the input is unity is shown in Fig. 6-4-1. As discussed later, when $\omega = 0$ the typical output is of unit magnitude. Much of the information contained in this curve can be summarized by 3 terms:

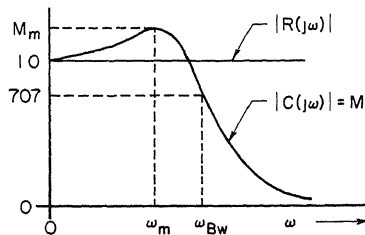


Fig. 6-4-1.

1. The maximum value reached by the magnitude of the output is denoted by M_m .
2. The value of ω at which the output reaches its maximum value M_m is denoted by ω_m .
3. The value of ω at which the output reaches 0.707 of the input is called the bandwidth and is denoted by ω_{BW} .

6-5. Block Diagram Transformations. In Chapters 3, 4, and 5, the transfer function and its associated block diagram are introduced. Since the block diagram method of analyzing control systems is extremely useful, the manipulation of more elaborate diagrams is helpful. By the use of transformation rules, a complicated diagram can be reduced to a single block. Several of these rules are developed in this article and a few of those developed in Chapter 3 are reviewed.

Two transfer functions connected in tandem can be multiplied to yield the transfer function of the combination. This concept is discussed

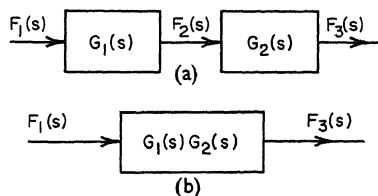


Fig. 6-5-1.

in Chapter 3 and is shown in Fig. 6-5-1 for review. The arrows show the direction of signal flow. The transfer function is defined as the ratio of the Laplace transform of the output to the Laplace of the input. In Fig. 6-5-1a, the signal $F_2(s)$ can be written as

$$F_2(s) = G_1(s)F_1(s) \quad (6-5-1)$$

The signal $F_3(s)$ is

$$F_3(s) = G_2(s)F_2(s) = G_1(s)G_2(s)F_1(s) \quad (6-5-2)$$

In Fig. 6-5-1b, the signal $F_3(s)$ can be written as

$$F_3(s) = G_1(s)G_2(s)F_1(s) \quad (6-5-3)$$

a comparison of Eqs. 6-5-2 and 6-5-3 shows the two block diagrams to be equivalent.

A summing point is a point in a system where two or more signals combine. Fig. 6-5-2a indicates that the signals $F_1(s)$ and $F_2(s)$ combine

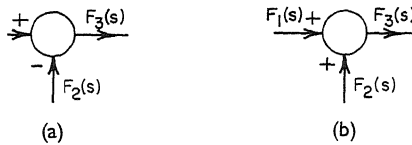


Fig. 6-5-2.

to yield $F_3(s)$. The plus sign for $F_1(s)$ indicates that $F_1(s)$ goes through the summing point with a plus sign whereas the minus sign on $F_2(s)$ indicates that $F_2(s)$ goes through the summing point with a minus sign as given by the equation

$$F_1(s) - F_2(s) = F_3(s) \quad (6-5-4)$$

Fig. 6-5-2b shows two plus signs indicating the signals combine as

$$F_1(s) + F_2(s) = F_3(s) \quad (6-5-5)$$

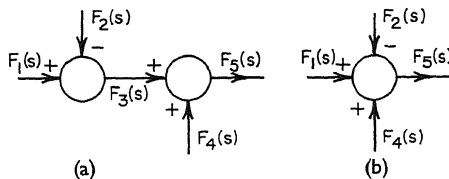


Fig. 6-5-3.

Summing points in tandem can be combined as indicated in Fig. 6-5-3. For the system of Fig. 6-5-3a, the output of the first summing point $F_3(s)$ is given by

$$F_3(s) = F_1(s) - F_2(s) \quad (6-5-6)$$

The output of the second summing point is

$$F_5(s) = F_3(s) + F_4(s) = F_1(s) - F_2(s) + F_4(s) \quad (6-5-7)$$

The output of the system of Fig. 6-5-3b is also given by this equation and therefore the systems of Fig. 6-5-3a and 6-5-3b are equivalent.

In certain situations, it may be convenient to divide one summing point with a number of incoming signals into two or more summing points in tandem. Fig. 6-5-3 demonstrates this when considered in the reverse direction.

A pick-off point is a point in a system at which the signal proceeds along several different paths without being affected by the presence of these paths. A pick-off point is indicated in Fig. 6-5-4.

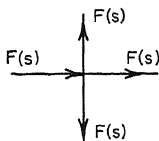


Fig. 6-5-4.

The basic block diagram for a closed-loop system is shown in Fig. 6-5-5a and the equivalent single block that replaces the closed-loop

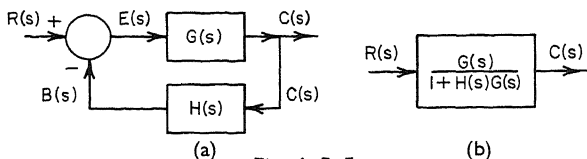


Fig. 6-5-5.

system is shown in Fig. 6-5-5b. The equivalence of these two systems is shown in the following development. In Fig. 6-5-5a, $G(s)$ is the transfer function of the forward portion of the system and $H(s)$ is the transfer function of the feedback portion of the system. In reality, as discussed in Art. 6-7 these blocks may be a combination of many sub-blocks which are necessary in analyzing the original system.

The input signal to the closed-loop system is $r(t)$ and its Laplace transform is $R(s)$. The output signal is $c(t)$ which has a transform $C(s)$. The actuating signal and its transform are $e(t)$ and $E(s)$ respectively. The primary feedback signal and its transform are $b(t)$ and $B(s)$ respectively.

The equations that describe the system are

$$\begin{aligned} E(s) &= R(s) - B(s) \\ C(s) &= G(s)E(s) \\ B(s) &= H(s)C(s) \end{aligned} \quad (6-5-8)$$

The first of these equations describes the summing point; the others result from the definition of a transfer function. The equivalent transfer function of Fig. 6-5-5b is obtained by eliminating $E(s)$ and $B(s)$ from these equations. The first two equations are combined as

$$C(s) = G(s)[R(s) - B(s)] \quad (6-5-9)$$

and $B(s)$ in the resulting equation is replaced by the third equation

$$C(s) = G(s)[R(s) - H(s)C(s)] \quad (6-5-10)$$

This equation can be rewritten as

$$C(s)[1 + H(s)G(s)] = G(s)R(s)$$

which can be solved for the transfer function of the closed-loop system as

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + H(s)G(s)} \quad (6-5-11)$$

If $B(s)$ goes through the summing point with a plus sign instead of the minus sign of Fig. 6-5-5a, the transfer function for the closed-loop system becomes

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 - H(s)G(s)} \quad (6-5-12)$$

This transfer function should be compared with the corresponding well-known equation from feedback amplifier theory

$$\frac{V_{out}}{V_{in}} = \frac{K}{1 - \beta K}$$

To obtain this equation, the voltage feedback in the amplifier passes through the summing point with a plus sign.

In order to derive the transfer function $E(s)/R(s)$, $C(s)$ in Eq. 6-5-11 can be replaced by $G(s)E(s)$ and the $G(s)$ canceled in both sides of the equation, resulting in

$$\frac{E(s)}{R(s)} = \frac{1}{1 + H(s)G(s)} \quad (6-5-13)$$

Eqs. 6-5-11 and 6-5-13 should be memorized as they are used often in the remainder of the book.

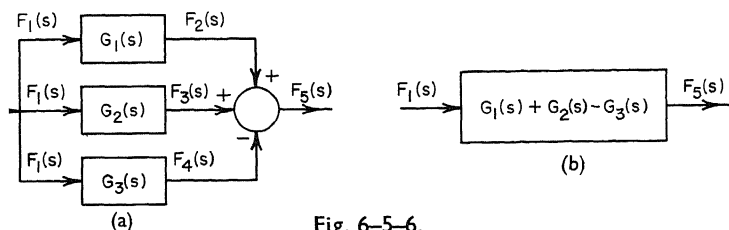


Fig. 6-5-6.

A group of transfer functions connected in parallel can be combined as shown in Fig. 6-5-6. For the system of Fig. 6-5-6a, the output of the summing points is

$$F_2(s) + F_3(s) - F_4(s) = F_5(s) \quad (6-5-14)$$

This equation can be rewritten as

$$F_1(s)G_1(s) + F_1(s)G_2(s) - F_1(s)G_3(s) = F_5(s) \quad (6-5-15)$$

or as

$$F_1(s)[G_1(s) + G_2(s) - G_3(s)] = F_5(s) \quad (6-5-16)$$

Since the system of Fig. 6-5-6b is also described by the equation, the systems of Fig. 6-5-6a and b are equivalent systems.

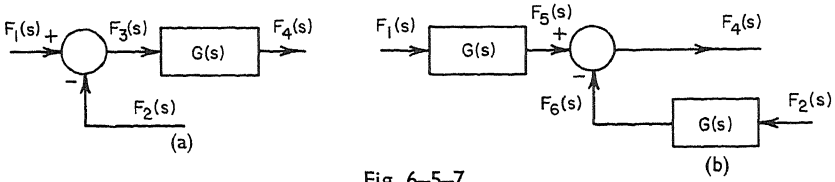


Fig. 6-5-7.

A transfer function block can be moved ahead of a summing point as shown in Fig. 6-5-7. For the system of Fig. 6-5-7a, the output of the summing point is

$$F_3(s) = F_1(s) - F_2(s) \quad (6-5-17)$$

The output of the entire system is

$$F_4(s) = G(s)F_3(s) = G(s)F_1(s) - G(s)F_2(s) \quad (6-5-18)$$

For the system of Fig. 6-5-7b, the output of the summing point is

$$F_4(s) = F_5(s) - F_6(s) \quad (6-5-19)$$

which can be written as

$$F_4(s) = G(s)F_1(s) - G(s)F_2(s) \quad (6-5-20)$$

Since Eqs. 6-5-18 and 6-5-20 are identical, the systems of Fig. 6-5-7a and b are equivalent.

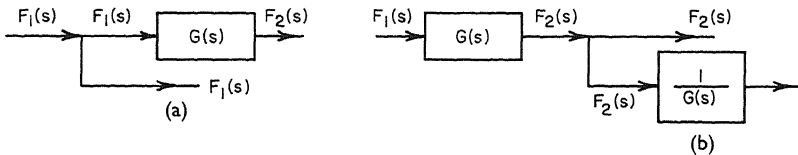


Fig. 6-5-8.

A transfer function block can be moved ahead of a pick-off point as shown in Fig. 6-5-8. The proof of this procedure is left as Prob. 6-3.

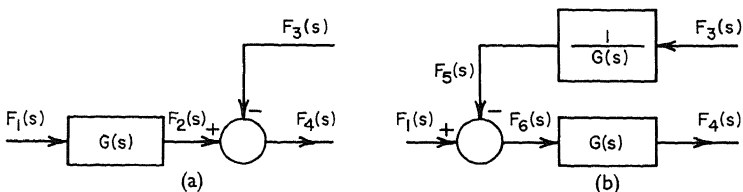


Fig. 6-5-9.

A transfer function block can be moved behind a summing point as shown in Fig. 6-5-9. The proof of this procedure is left as Prob. 6-4.

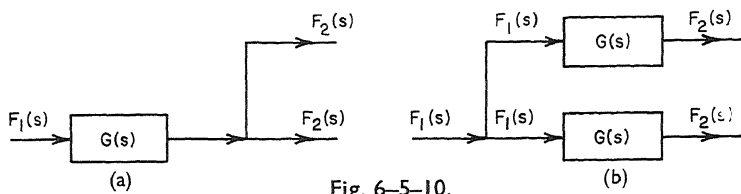


Fig. 6-5-10.

A transfer function block can be moved behind a pick-off point as shown in Fig. 6-5-10. The proof of this procedure is left as Prob. 6-5.

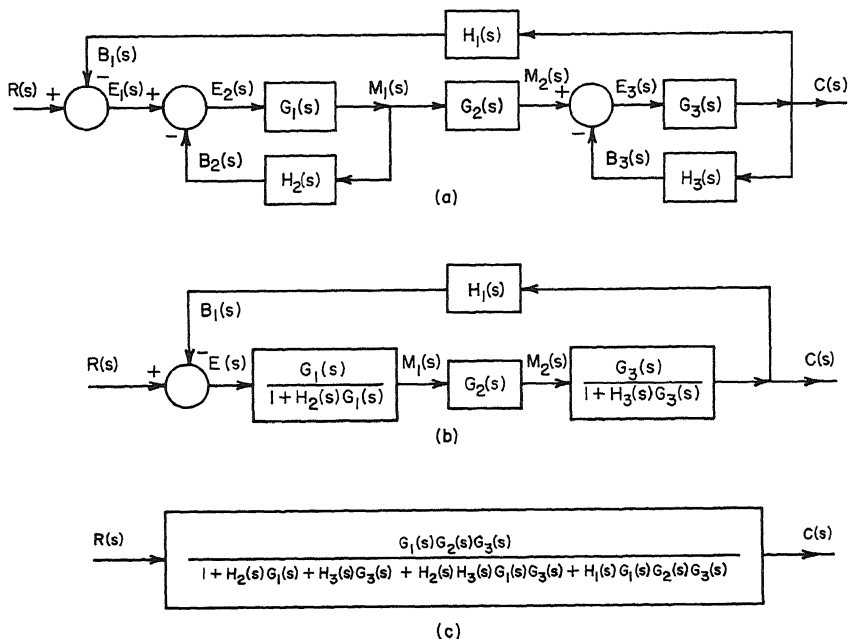


Fig. 6-6-1.

6-6. Examples Using Block Diagram Transformations. The following examples demonstrate the procedure of the previous article.

Example 6-6-1. The system of Fig. 6-6-1a is to be reduced to a single block. Each of the two minor loops can be replaced by a single block as shown in Fig. 6-6-1b. The three blocks in the forward portion of the system can be combined into a single $G(s)$ as

$$G(s) = \frac{G_1(s)}{1 + H_2(s)G_1(s)} \cdot G_2(s) \cdot \frac{G_3(s)}{1 + H_3(s)G_3(s)} \quad (6-6-1)$$

The closed-loop of Fig. 6-6-1b can be replaced by a single block which upon simplification yields the results shown in Fig. 6-6-1c.

Example 6-6-2. The system of Fig. 6-6-2a is to be reduced to a single block. The systems of Fig. 6-6-2a is transformed into the system of Fig. 6-6-2b. The $G_3(s)$ block is moved ahead of the pick-off point and the $G_1(s)$ block is moved behind the summing point. The equivalent transfer function $G(s)$ for the forward loop of Fig. 6-6-2b is

$$G(s) = G_1(s)G_2(s)G_3(s) \quad (6-6-2)$$

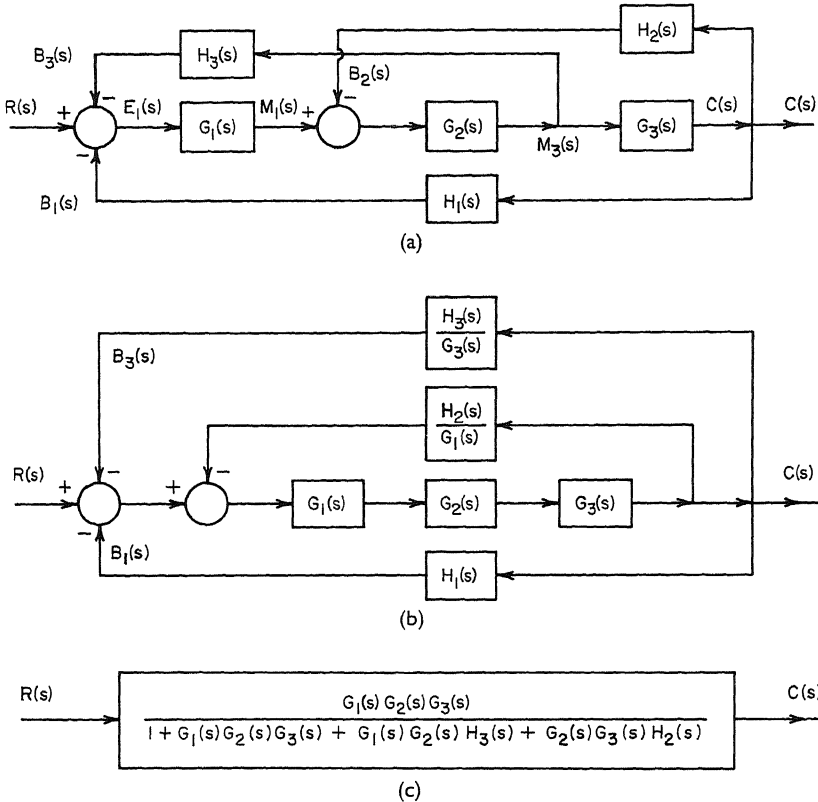


Fig. 6-6-2.

The three elements in the feedback path are in parallel and the equivalent $H(s)$ is

$$H(s) = H_1(s) + \frac{H_2(s)}{G_1(s)} + \frac{H_3(s)}{G_3(s)} \quad (6-6-3)$$

When $H(s)$ and $G(s)$ are combined and simplified, the resulting single transfer function is given in Fig. 6-6-2c.

It is interesting to compare the final result of Fig. 6-6-2c with the original system of Fig. 6-6-2a. The forward portion is made up of the blocks $G_1(s)$, $G_2(s)$ and $G_3(s)$ and the product of these three transfer functions appears in the numerator of Fig. 6-6-2c. The denominator is made up of 1 plus the sum of the transfer functions around the various loops. The product $G_1(s)G_2(s)G_3(s)H_1(s)$ can be identified with one loop, the product $G_1(s)G_2(s)H_3(s)$ with a second loop and the product of $G_2(s)G_3(s)H_2(s)$ with the third loop.

The concepts discussed in the last paragraph cannot be pressed too far as can be seen by inspection of the system of Fig. 6-6-1 which contains an extra term in the denominator of Fig. 6-6-1c as compared with the number of loops in Fig. 6-6-1a. A discussion of why the extra term appears in one system and not in the other is left as Prob. 6-14.

6-7. Obtaining Block Diagrams for Simple Systems. Several of the following articles analyze simple systems based on a block diagram. Although a certain nomenclature is used, the block diagram can represent any one of many systems. In this article two systems are reduced to block diagrams of the same form to indicate the method.

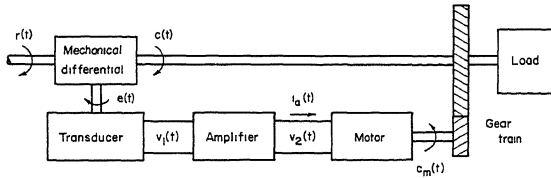


Fig. 6-7-1.

Example 6-7-1. The rotational mechanical system shown schematically in Fig. 6-7-1 is the first example.

The mechanical differential can be replaced by a synchro system if the input shaft $r(t)$ and output shaft $c(t)$ are at remote locations. For present purposes the mechanical differential is essentially a summing point defined by

$$e(t) = r(t) - c(t) \quad (6-7-1)$$

The transducer is any device yielding an output voltage which depends in magnitude and sense upon the rotation of the $e(t)$ shaft. One such device is that of

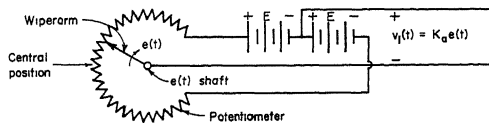


Fig. 6-7-2.

Fig. 6-7-2. The $e(t)$ shaft moves the wiper arm on a potentiometer. Two equal d-c voltages of magnitude E are placed across the potentiometer with the center tap brought out as one of the two output leads. The other output lead is connected to the wiper arm. If the $e(t)$ shaft is in its neutral position, the wiper arm is at its central position and the output voltage $v_1(t)$ is zero. If $e(t)$ is turned in the positive or negative direction, $v_1(t)$ is proportional to $e(t)$ and has the same sign as given by

$$v_1(t) = K_a e(t) \quad (6-7-2)$$

Some potentiometers are made in the form of a helix so that the $e(t)$ shaft can rotate through several turns. An analysis of the coefficient K_a in terms of quantities shown in Fig. 6-7-2 is left as Prob. 6-46.

Since the usual frequency range of a control system is limited to relatively low frequencies, the d-c amplifier can be assumed to have a constant gain as given by

$$v_2(t) = K_b v_1(t) \quad (6-7-3)$$

The voltage $v_2(t)$ is applied to the armature of a d-c motor which has a separately excited field. This analysis is similar to that of Chapter 4 Art. 4-11. In this analysis, the time constants of the electrical portion of the system are neglected. In many cases this is possible because the time constants of the mechanical portion of the system are so much larger than the electrical time constants. Therefore the total response of the system is determined almost entirely by its mechanical portion. If in a particular situation the electrical time constants cannot be ignored, they can be added in a straightforward manner although the order of the differential equation is increased.

In this analysis, the voltage $v_2(t)$ is applied to a motor with an armature current i_a , an armature resistance R_a and a back emf constant K_{av} and these terms are related by

$$v_2(t) = R_a i_a(t) + K_{av} \frac{dc_m(t)}{dt} \quad (6-7-4)$$

The torque constant of the motor is K_T and the torque output $t(t)$ of the motor is given by

$$t(t) = K_T i_a(t) \quad (6-7-5)$$

This torque is applied to the rotational system. In terms of the motor shaft, all of the rotational moments of inertia such as armature of the motor, shafts, gear train and the load are labelled J_m . Similarly, all the components of viscous friction are added and in terms of the motor shaft are labelled D_m . The equation relating the torque $t(t)$ and the rotation system is

$$t(t) = J_m \frac{d^2 c_m(t)}{dt^2} + D_m \frac{dc_m(t)}{dt} \quad (6-7-6)$$

Finally the gear train is such that the motor shaft $c_m(t)$ rotates n times for each revolution of the output shaft $c(t)$.

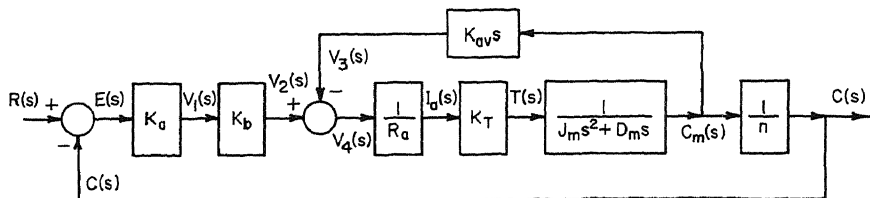


Fig. 6-7-3.

The preceding equations are transformed and the system block diagram is indicated in Fig. 6-7-3. The transforms of Eq. 6-7-2 and 6-7-3 yield the blocks labeled K_a and K_b respectively.

The summing point with input voltages $V_2(s)$ and $V_3(s)$ needs to be examined. In Eq. 6-7-4, the voltage component $R_a i_a(t)$ is labeled $v_4(t)$, the component $K_{av} dc_m(t)/dt$ is labeled $v_3(t)$ and the equation can be written as

$$v_4(t) = v_2(t) - v_3(t) \quad (6-7-7)$$

The transform of this equation describes the summing point.

The voltage equation

$$v_3(t) = K_{av} \frac{dc_m(t)}{dt} \quad (6-7-8)$$

is transformed as

$$V_3(s) = K_{av} s C_m(s) \quad (6-7-9)$$

and is represented by a signal being fed back from the $C_m(s)$ shaft.

The voltage equation

$$v_a(t) = R_a i_a(t) \quad (6-7-10)$$

transforms to yield the $1/R_a$ block and Eq. 6-7-5 yields the K_T block. Eq. 6-7-6 repeated as

$$i(t) = J_m \frac{d^2 c_m(t)}{dt^2} + D_m \frac{dc_m(t)}{dt} \quad (6-7-11)$$

transforms as

$$T(s) = [J_m s^2 + D_m s] C_m(s) \quad (6-7-12)$$

and the block relating $C_m(s)$ to $T(s)$ results.

Finally, the action of the gear train produces the $1/n$ block.

The minor-loop can be replaced by a single transfer function

$$\frac{K_T / (R_a [J_m s^2 + D_m s])}{1 + \frac{K_{av} K_T s}{R_a [J_m s^2 + D_m s]}} = \frac{K_T / R_a}{J_m s^2 + \left(D_m + \frac{K_{av} K_T}{R_a} \right) s} \quad (6-7-13)$$

When this transfer function is multiplied by the other transfer functions of the forward loop, the total $G(s)$ becomes

$$G(s) = \frac{\left(\frac{K_a K_b K_T}{R_a} \right) n}{J_m n^2 s^2 + \left(D_m + \frac{K_{av} K_T}{R_a} \right) n^2 s} \quad (6-7-14)$$

The numerator and denominator of this $G(s)$ are multiplied by n to aid in the following discussion.

If the $C(s)$ shaft is blocked and the $E(s)$ shaft turned through one radian, the factor $K_a K_b K_T / R_a$ represents the torque output of the motor in terms of the motor shaft. When this factor is multiplied by n , the torque is given in terms of the output shaft and this factor is replaced by K .

$$K = \left(\frac{K_a K_b K_T}{R_a} \right) n \quad (6-7-15)$$

The term J_m is the total moment of inertia expressed in terms of the motor shaft. When this term is multiplied by n^2 , the total moment of inertia expressed in terms of the output shaft results and is denoted by

$$J = J_m n^2 \quad (6-7-16)$$

Examination of the term

$$D_m + \frac{K_{av} K_T}{R_a} \quad (6-7-17)$$

shows that the back emf of the motor acting around the minor loop effectively increases the magnitude of viscous friction. When this term is multiplied by n^2 , the effective viscous friction in terms of the output shaft is obtained and is called

$$D = \left(D_m + \frac{K_{av} K_T}{R_a} \right) n^2 \quad (6-7-18)$$

With the introduction of the new nomenclature, the transfer function of the forward loop becomes

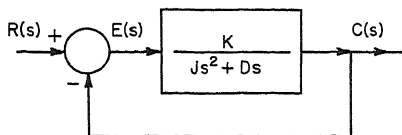


Fig. 6-7-4.

$$G(s) = \frac{K}{Js^2 + Ds} \quad (6-7-19)$$

and the equivalent block diagram for the closed-loop system is shown in Fig. 6-7-4.

Example 6-7-2. A closed-loop control system of a different type is shown in Fig. 6-7-5. In this system a hydraulic pilot valve controls an actuator piston which

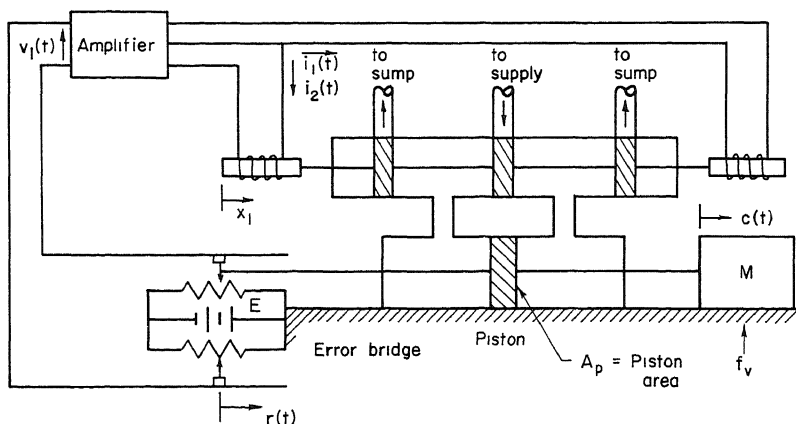


Fig. 6-7-5.

positions a mass M whose motion is opposed by viscous friction, f_v . The piston also moves a wiper arm of a linear motion potentiometer on one side of an error measuring bridge. The input position is applied to the wiper arm of the potentiometer on the other side of the error bridge. When the input and output positions are the same no error voltage exists and there is no input to the error amplifier. When an error voltage, $v_1(t)$, exists the amplifier causes one of the output currents $i_1(t)$ or $i_2(t)$ to increase and the other to decrease. This current difference causes the pilot valve to open and the high pressure fluid moves the mass and tends to rebalance the bridge.

The analysis of the valve-piston combination parallels that of Chapter 5; however, the equations are repeated here in order to allow easy comparison with the block diagram of Fig. 6-7-6. The output voltage of the error bridge, $v_1(t)$, is related to the position error, $e(t)$, by

$$v_1(t) = K_a e(t) \quad (6-7-20)$$

Similarly the amplifier is assumed to produce a linear proportional relation between its input, $v_1(t)$, and the output difference current, $\Delta i(t) = i_1(t) - i_2(t)$,

$$\Delta i(t) = K_b v_1(t) \quad (6-7-21)$$

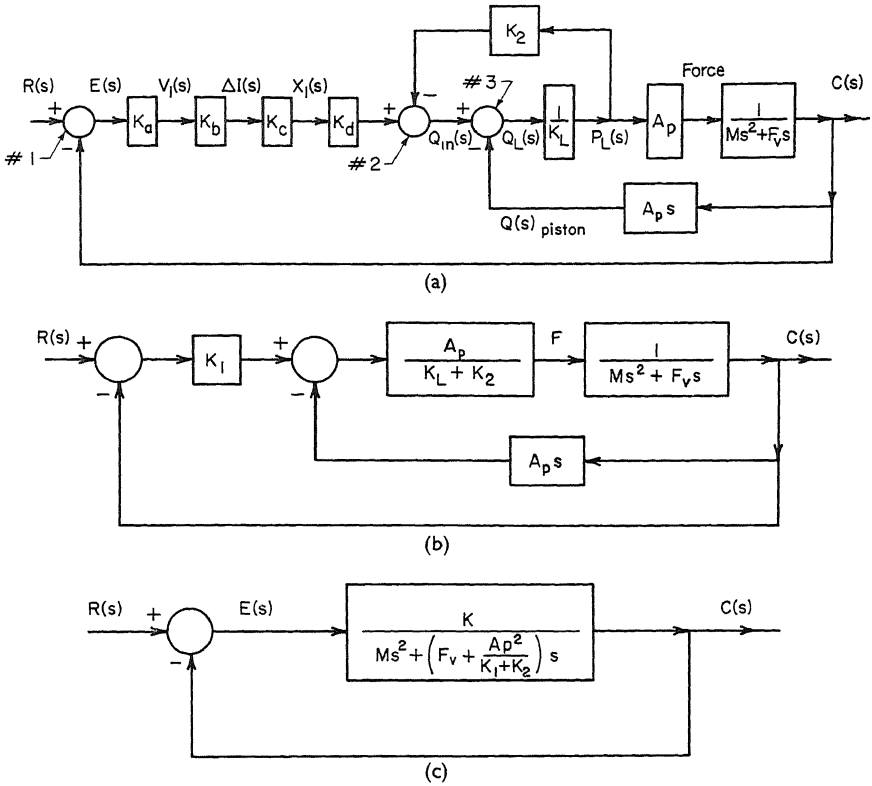


Fig. 6-7-6.

This equation ignores the solenoid inductances. As a next step, the valve movement, $x_1(t)$, is assumed to be proportional to $\Delta i(t)$,

$$x_1(t) = K_c \Delta i(t) \quad (6-7-22)$$

The actual input flow rate is established by using the incremental approach of Chapter 5 as follows. The input flow rate depends on the valve opening, $x_1(t)$, and on the load pressure drop across the actuator piston, $p_L(t)$. For small variations as discussed in Chapter 5

$$q_{in}(t) = K_d x_1(t) - K_2 p_L(t) \quad (6-7-23)$$

Eq. 6-7-23 leads to summing point 2 in Fig. 6-7-6a. The input flow rate is equal to the load flow rate to the actuator piston, $A_p dc(t)/dt$, plus the leakage flow, $K_L p_L$. This relationship is depicted at summing point 3 in fig. 6-7-6a.

$$q_{in} - q_{\text{piston}} = q_{\text{leak}} \quad (6-7-24)$$

$$q_{in} - A_p \frac{dc(t)}{dt} = K_L p_L \quad (6-7-25)$$

The remaining equations result from Newton's second law.

$$\text{Force} = A_p p_L(t) = M \frac{d^2 c(t)}{dt^2} + f \frac{dc(t)}{dt} \quad (6-7-26)$$

The transforms of Eqs. 6-7-20 through 6-7-26 allow the determination of all blocks in Fig. 6-7-6a. The problem then becomes one of reducing blocks. The first step is that of combining gains K_a , K_b , K_c , K_d into K_1 . Next, summing points 2 and 3 are combined and the minor loop consisting of $1/K_L$ and K_2 is reduced to $1/(K_L + K_2)$ by using $G(s)/[1 + G(s)H(s)]$. Fig. 6-7-6b shows the results of this reduction. The last step requires reducing the remaining minor loop to give Fig. 6-7-6c in which the substitution

$$K = \frac{K_1 A_p}{K_L + K_2} \quad (6-7-27)$$

has been made. Thus the hydraulic system of this example can be reduced to a block diagram of form exactly like the armature controlled motor of Ex. 6-7-1.

6-8. Transient Analysis of a Simple System. The closed-loop system to be analyzed is that of Fig. 6-7-4. The closed-loop transfer function is

$$\begin{aligned} \frac{C(s)}{R(s)} &= \frac{G(s)}{1 + G(s)} = \frac{K/(Js^2 + Ds)}{1 + K/(Js^2 + Ds)} \\ &= \frac{K}{Js^2 + Ds + K} = \frac{K/J}{s^2 + (D/J)s + K/J} \end{aligned} \quad (6-8-1)$$

As discussed in the previous article, this form of the block diagram and the resulting closed-loop transfer function can represent one of many systems even though the nomenclature used is that of a rotational mechanical system. The following analysis is completely general for any system which has a closed-loop transfer function containing two poles and no finite zeros.

Since the system has unity feedback, the actuating signal becomes the error signal. The transfer function which is the ratio of $E(s)$ to $R(s)$ is given by

$$\frac{E(s)}{R(s)} = \frac{1}{1 + G(s)} = \frac{s^2 + (D/J)s}{s^2 + (D/J)s + K/J} \quad (6-8-2)$$

In the following articles, the output and error functions are investigated for various inputs.

6-9. Unit-Impulse Input. The first input to be considered is the unit-impulse function. It may seem unrealistic to consider applying an impulse to a physical system as the concept implies an infinite magnitude for zero time. However, as a practical matter a finite input R can be applied for a short length of time T_1 . The time T_1 must be chosen so that it is small compared with the response time of the system. A method of testing to determine whether T_1 is sufficiently small is to apply an input $2R$ for a time $T_1/2$. If the original and the new response are essentially the same, T_1 is sufficiently small.

For present purposes, this type of input is used because of the convenience offered by such a function. When $r(t) = u_0(t)$, $R(s) = 1$, and $C(s)$ can be found from Eq. 6-8-1 as

$$C(s) = \frac{K/J}{s^2 + (D/J)s + K/J} \quad (6-9-1)$$

The system response may be oscillatory, critically damped or overdamped depending on the relationship among D , J and K . For the first portion of this analysis, the oscillatory case is assumed and the denominator is written as

$$C(s) = \frac{K/J}{(s + D/2J)^2 + \left(\sqrt{K/J - D^2/4J^2}\right)^2} \quad (6-9-2)$$

For convenience in writing, two letters are defined as

$$a = \frac{D}{2J} \quad \text{and} \quad \omega = \sqrt{\frac{K}{J} - \frac{D^2}{4J^2}} \quad (6-9-3)$$

and $C(s)$ becomes

$$C(s) = \frac{K/J}{(s + a)^2 + \omega^2} \quad (6-9-4)$$

Item 13 in the Transform Table in the Appendix is used as

$$A/a = \frac{K}{J\omega} \quad (6-9-5)$$

yielding

$$c(t) = \frac{K}{J\omega} e^{-at} \sin \omega t \quad (6-9-6)$$

The condition that the system is oscillatory demands that

$$\frac{D^2}{4J^2} < \frac{K}{J}$$

or

$$D < 2\sqrt{KJ} \quad (6-9-7)$$

For analytical purposes, K and J are considered fixed and D is varied from zero and increased until the overdamped case is reached.

When D is equal to zero, the system oscillates at an ω which is labeled ω_n and is called the undamped natural radian frequency of the system as given by

$$\omega_n = \sqrt{\frac{K}{J}} \quad (6-9-8)$$

Under these conditions, the equation for the output is

$$c(t) = \frac{K}{J\omega_n} \sin \omega_n t = \omega_n \sin \omega_n t \quad (6-9-9)$$

As D increases, Eq. 6-9-6 describes the output until D reaches a value D_c called critical damping as given by

$$D_c = 2\sqrt{KJ} \quad (6-9-10)$$

For this critically damped case ω is equal to zero and Eq. 6-9-6 becomes

$$c(t) = \omega_n^2 e^{-at} \lim_{\omega \rightarrow 0} \frac{\sin \omega t}{\omega} = \omega_n^2 t e^{-at} \quad (6-9-11)$$

For D greater than D_c , ω becomes imaginary and a letter b is defined as

$$b = \sqrt{\frac{D^2}{4J^2} - \frac{K}{J}} \quad (6-9-12)$$

so that

$$\omega = jb \quad (6-9-13)$$

Under these conditions, the output as obtained from Eq. 6-9-6 becomes

$$c(t) = \frac{\omega_n^2}{jb} \varepsilon^{-at} \sin jbt = \frac{\omega_n^2}{b} \varepsilon^{-at} \sinh bt \quad (6-9-14)$$

6-10. Generalized Equations. An examination of the system equations developed in the previous article indicates that each equation can be expressed in more general form in terms of ω_n and the ratio of the damping present to the value of critical damping. By definition the letter ζ is used to express this ratio,

$$\zeta = \frac{D}{D_c} \quad (6-10-1)$$

The other coefficients can be expressed in terms of ζ and ω_n as

$$a = \frac{D}{2J} = \frac{2\zeta\sqrt{KJ}}{2J} = \zeta\sqrt{\frac{K}{J}} = \zeta\omega_n \quad (6-10-2)$$

$$\omega = \sqrt{\frac{K}{J} - \frac{D^2}{4J^2}} = \sqrt{\omega_n^2 - \zeta^2\omega_n^2} = \omega_n\sqrt{1 - \zeta^2} \quad (6-10-3)$$

$$b = \sqrt{\frac{D^2}{4J^2} - \frac{K}{J}} = \omega_n\sqrt{\zeta^2 - 1} \quad (6-10-4)$$

The equation for the oscillatory system as developed in the previous article for the response of the system to a unit-impulse can be written as

$$c(t) = \frac{K}{J\omega} \varepsilon^{-at} \sin \omega t = \frac{\omega_n}{\sqrt{1 - \zeta^2}} \varepsilon^{-\zeta\omega_n t} \sin \sqrt{1 - \zeta^2} \omega_n t \quad (6-10-5)$$

This equation is not yet in the best form for plotting as a different curve results for each value of ω_n . However, if $c(t)$ is divided by ω_n and the ratio is plotted versus $\omega_n t$, only one curve results for each value of ζ . The final forms for the equations are

Oscillatory Case

$$\frac{c(t)}{\omega_n} = \frac{1}{\sqrt{1 - \zeta^2}} \varepsilon^{-\zeta\omega_n t} \sin \sqrt{1 - \zeta^2} \omega_n t \quad (6-10-6)$$

Critically Damped Case

$$\frac{c(t)}{\omega_n} = \omega_n t \varepsilon^{-\omega_n t} \quad (6-10-7)$$

Overdamped Case

$$\frac{c(t)}{\omega_n} = \frac{1}{\sqrt{\zeta^2 - 1}} e^{-\zeta \omega_n t} \sinh \sqrt{\zeta^2 - 1} \omega_n t \quad (6-10-8)$$

This family of curves is plotted in Fig. 6-10-1.

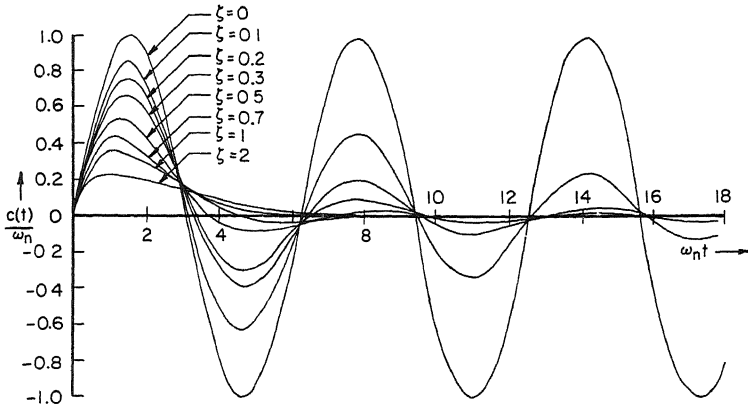


Fig. 6-10-1.

The equations for $e(t)$ can be determined from $E(s)$ which in turn can be found from Eq. 6-8-2 by letting $R(s) = 1$

$$E(s) = \frac{s^2 + (D/J)s}{s^2 + (D/J)s + K/J} = 1 - \frac{K/J}{s^2 + (D/J)s + K/J} \quad (6-10-9)$$

This equation displays the fact that $e(t)$ has an impulse at $t = 0$ and is thereafter the negative of $c(t)$. This result is apparent from

$$e(t) = r(t) - c(t) = u_0(t) - c(t) \quad (6-10-10)$$

Generalized curves of $e(t)/\omega_n$ are shown in Fig. 6-10-2.

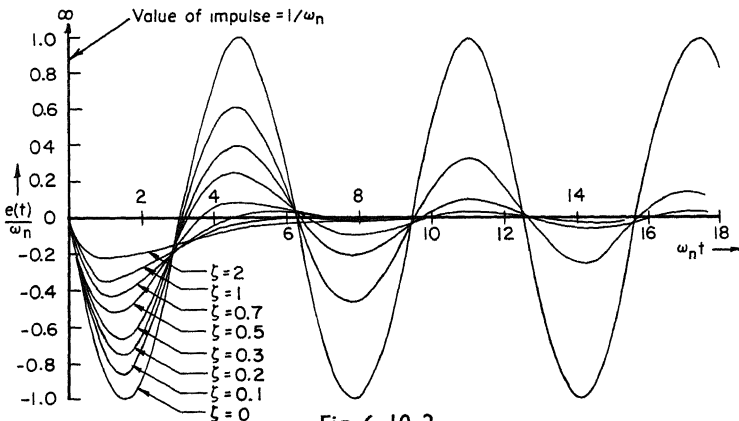


Fig. 6-10-2.

6-11. Unit-Step Function Input. The response of a system to a unit-step function is the integral of the response of the system to a unit-impulse function. However, for the present purposes, Eqs. 6-8-1 and 6-8-2 are solved by $C(s)$ and $E(s)$ respectively with $R(s) = 1/s$.

$$C(s) = \frac{K/J}{s[s^2 + (D/J)s + K/J]} \quad (6-11-1)$$

$$E(s) = \frac{s^2 + (D/J)s}{s[s^2 + (D/J)s + K/J]} \quad (6-11-2)$$

These equations can be expressed in terms of ω_n and ζ as

$$C(s) = \frac{\omega_n^2}{s[s^2 + 2\zeta\omega_n s + \omega_n^2]} \quad (6-11-3)$$

$$\begin{aligned} E(s) &= \frac{s^2 + 2\zeta\omega_n s}{s[s^2 + 2\zeta\omega_n s + \omega_n^2]} \quad (6-11-4) \\ &= \frac{1}{s} - \frac{\omega_n^2}{s[s^2 + 2\zeta\omega_n s + \omega_n^2]} = \frac{1}{s} - C(s) \end{aligned}$$

As soon as $c(t)$ is determined, $e(t)$ can be written as

$$e(t) = u_{-1}(t) - c(t) \quad (6-11-5)$$

Item 19 in the Transform Table in the Appendix is used to find $c(t)$ as

$$c(t) = 1 + \frac{1}{\sqrt{1 - \zeta^2}} \varepsilon^{-\zeta\omega_n t} \sin \left[\sqrt{1 - \zeta^2} \omega_n t - \tan^{-1} \frac{\sqrt{1 - \zeta^2}}{-\zeta} \right] \quad (6-11-6)$$

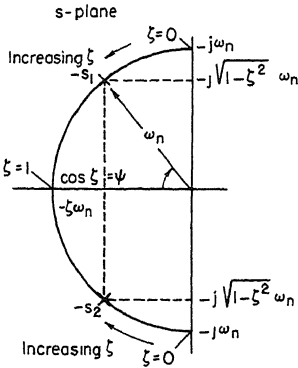


Fig. 6-11-1.

This equation assumes that $0 \leq \zeta < 1$. The pole-zero pattern for $C(s)$ is pictured in Fig. 6-11-1. Except for a multiplying constant, the pole-zero configuration completely describes the system's behavior. This concept is discussed in the following paragraphs and at several other points in the book. The pole at the origin results from $R(s) = 1/s$. When $\zeta = 0$, the poles originating in the closed-loop transfer function are on the $j\omega$ axis at $\pm j\omega_n$.

The product of the two poles is equal to ω_n^2 and is independent of ζ ; therefore as ζ increases, these conjugate poles move along the circular path of radius ω_n .

The poles are located at

$$-s_1 \text{ and } -s_2 = -\zeta\omega_n \pm j\sqrt{1 - \zeta^2}\omega_n \quad (6-11-7)$$

and a pair of poles is indicated in Fig. 6-11-1. The angle from the negative real axis to the radial line from the origin to the pole at $-s_1$, is defined as ψ . $\cos \psi$ is given by

$$\cos \psi = \frac{\zeta \omega_n}{\omega_n} = \zeta \quad (6-11-8)$$

From this $\sin \psi$ and $\tan \psi$ are found as

$$\sin \psi = \sqrt{1 - \cos^2 \psi} = \sqrt{1 - \zeta^2} \quad (6-11-9)$$

$$\tan \psi = \frac{\sqrt{1 - \zeta^2}}{\zeta} \quad (6-11-10)$$

With these identities, Eq. 6-11-6 can be written as

Oscillatory Case

$$c(t) = 1 - \frac{\varepsilon^{-\zeta \omega_n t}}{\sin \psi} \sin [\omega_n t \sin \psi + \psi] \quad (6-11-11)$$

where $0 \leq \zeta < 1$.

When $\zeta = 1$, the two poles meet on the negative real axis and Eq. 6-11-3 can be written as

$$C(s) = \frac{\omega_n^2}{s(s + \omega_n)^2} \quad (6-11-12)$$

and $c(t)$ is

Critically Damped Case

$$c(t) = 1 - (1 + \omega_n t) \varepsilon^{-\omega_n t} \quad (6-11-13)$$

For $\zeta > 1$, the two poles stay on the negative real axis but separate, one going to the left and the other to the right so that their product remains ω_n^2 . The form for $c(t)$ becomes

Overdamped Case

$$c(t) = 1 - \varepsilon^{-\zeta \omega_n t} \left[\cosh \sqrt{\zeta^2 - 1} \omega_n t - \left(\frac{\zeta}{\zeta^2 - 1} \right) \sinh \sqrt{\zeta^2 - 1} \omega_n t \right] \quad (6-11-14)$$

Again if the equations for any of the three cases are plotted versus $\omega_n t$, only one curve results for each value of ζ . A family of such curves is given in Fig. 6-11-2. Because of Eq. 6-11-5 repeated as

$$e(t) = u_{-1}(t) - c(t) \quad (6-11-15)$$

the curves for $e(t)$ can be obtained from the $c(t)$ curve as also shown in Fig. 6-11-2.

One additional generalized curve available from the step function response for this system is the relationship between maximum overshoot and the damping ratio. Eq. 6-11-11 is differentiated with respect to time and this value of time is substituted into the equation yielding

$$c_{\max}(t) = 1 + e^{-\zeta\pi/\sqrt{1-\zeta^2}} \quad (6-11-16)$$

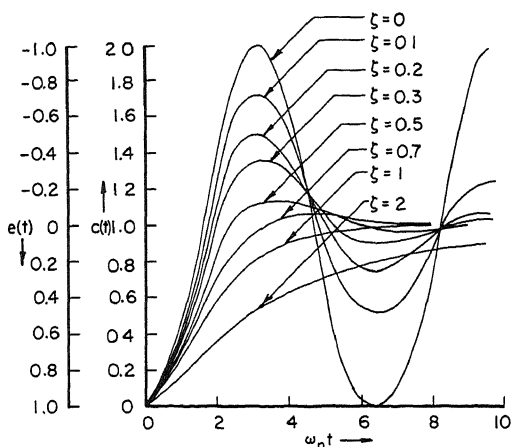


Fig. 6-11-2.

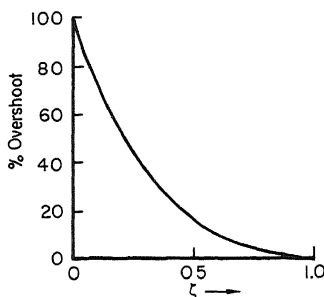


Fig. 6-11-3.

This equation can be rewritten to display the percent overshoot as

$$100 [c_{\max}(t) - 1] = 100 e^{-\zeta\pi/\sqrt{1-\zeta^2}} \quad (6-11-17)$$

The percent overshoot is plotted in Fig. 6-11-3.

6-12. Unit-Ramp Function Input. The response of a system to a unit-ramp function is the integral of the response of the system to a unit-step function. In order to use the Transform Table in the Appendix rather than integral tables, $C(s)$ and $E(s)$ from Eqs. 6-11-3 and 6-11-4 respectively are divided by s , giving

$$C(s) = \frac{\omega_n^2}{s^2[s^2 + 2\zeta\omega_n s + \omega_n^2]} \quad (6-12-1)$$

$$\begin{aligned} E(s) &= \frac{s^2 + 2\zeta\omega_n s}{s^2[s^2 + 2\zeta\omega_n s + \omega_n^2]} \\ &= \frac{1}{s^2} - \frac{\omega_n^2}{s^2[s^2 + 2\zeta\omega_n s + \omega_n^2]} = \frac{1}{s^2} - C(s) \end{aligned} \quad (6-12-2)$$

Item 26 in the Transform Table is used to find $c(t)$ as

$$c(t) = t - \frac{2\zeta}{\omega_n} + \frac{e^{-\zeta\omega_n t}}{\omega_n \sqrt{1-\zeta^2}} \sin \left[\sqrt{1-\zeta^2} \omega_n t + 2 \cos^{-1} \zeta \right] \quad (6-12-3)$$

where $0 \leq \zeta < 1$.

The error $e(t)$ is now

$$e(t) = \frac{2\zeta}{\omega_n} - \frac{\varepsilon^{-\zeta\omega_n t}}{\omega_n \sqrt{1-\zeta^2}} \sin [\sqrt{1-\zeta^2} \omega_n t + 2 \cos^{-1} \zeta] \quad (6-12-4)$$

where $0 \leq \zeta < 1$.

These equations must be rearranged slightly in order to be plotted in a generalized form.

Oscillatory Case

$$\omega_n c(t) = \omega_n t - 2\zeta + \frac{\varepsilon^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}} \sin [\sqrt{1-\zeta^2} \omega_n t + 2 \cos^{-1} \zeta] \quad (6-12-5)$$

$$\omega_n e(t) = 2\zeta - \frac{\varepsilon^{-\zeta\omega_n t}}{\sqrt{1-\zeta^2}} \sin [\sqrt{1-\zeta^2} \omega_n t + 2 \cos^{-1} \zeta] \quad (6-12-6)$$

Critically Damped Case

$$\omega_n c(t) = \omega_n t + \omega_n t \varepsilon^{-\omega_n t} - 2 + 2 \varepsilon^{-\omega_n t} \quad (6-12-7)$$

$$\omega_n e(t) = 2 - 2 \varepsilon^{-\omega_n t} - \omega_n t \varepsilon^{-\omega_n t} \quad (6-12-8)$$

Overdamped Case

$$\omega_n c(t) = \omega_n t - 2\zeta$$

$$+ \varepsilon^{-\zeta\omega_n t} \left[\frac{2\zeta^2 - 1}{\sqrt{\zeta^2 - 1}} \sinh \sqrt{\zeta^2 - 1} \omega_n t + 2\zeta \cosh \sqrt{\zeta^2 - 1} \omega_n t \right] \quad (6-12-9)$$

$$\omega_n e(t) = 2\zeta - \varepsilon^{-\zeta\omega_n t} \left[\frac{2\zeta^2 - 1}{\sqrt{\zeta^2 - 1}} \sinh \sqrt{\zeta^2 - 1} \omega_n t \right. \quad (6-12-10)$$

$$\left. + 2\zeta \cosh \sqrt{\zeta^2 - 1} \omega_n t \right]$$

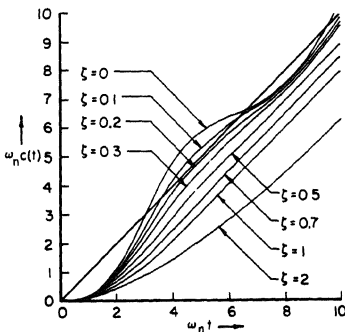


Fig. 6-12-1.

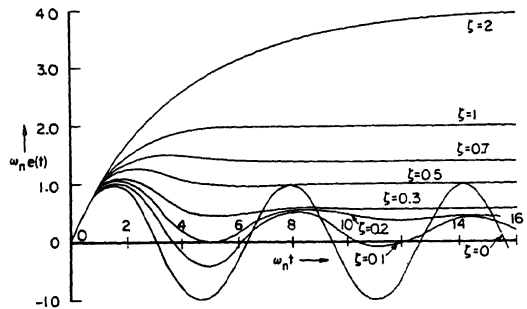


Fig. 6-12-2.

For various values of ζ , the family of curves for $c(t)$ is shown in Fig. 6-12-1 and for $e(t)$ in Fig. 6-12-2.

6-13. Discussion of the Set of Generalized Curves. The discussion in the previous article of the simple system has been presented in some detail because an understanding of this system is a key to understanding more complicated systems even though their response functions cannot be presented in terms of generalized curves. However, in many cases, certain aspects of the response of these systems can be deduced from the response curves for the simple system with two poles and no finite zeros. The discussion of the more complicated system is not attempted at this point; however, the generalized curves of the previous articles are reviewed with some of these later discussion in mind.

When $\zeta = 0$, the closed-loop poles are on the $j\omega$ -axis at $\pm j\omega_n$ and the system response oscillates indefinitely regardless of the driving function. This type of response is certainly undesirable because any disturbance will cause the system to oscillate indefinitely.

If D , the effective viscous friction, could become negative in a system of this sort, ζ would also be negative, and the complex poles would move into the right half s -plane. In this case the oscillations would build up until some portion of the system saturates and the system becomes nonlinear or in some cases until it is damaged. This type of system is said to be absolutely unstable. When D takes on a positive value, ζ is also positive and the poles move into the left half s -plane. In this case the oscillations to any type of input die away and the system is said to be absolutely stable.

The term relative stability is used to distinguish between two systems that are absolutely stable. The closer the complex conjugate pair of poles are to the $j\omega$ -axis, the more pronounced are the oscillations and the longer the time it takes for them to die away. To verify this statement, the system response to any of the input signals for $\zeta = 0.1$ and $\zeta = 0.7$ should be compared.

For the system being discussed, ζ is a measure of relative stability. It is extremely difficult to generalize about the desirable range of values for ζ because circumstances can alter cases so completely. However, in most cases, a system with a value of $\zeta = 2$ has no advantage over a system with $\zeta = 1$. For $\zeta = 2$, the system response to any input moves smoothly into its steady-state value with no oscillation. For $\zeta = 1$, the system response does likewise but does so more rapidly. Therefore in most cases a system with a $\zeta > 1$ offers no advantage over a system with a $\zeta = 1$. Furthermore, in a large majority of cases a certain amount of oscillations can be tolerated because the resulting system responds more rapidly. For these reasons an often-used "rule-of-thumb" is to design the system for a ζ approximately equal to 0.5. An immediate exception comes to mind. An automatically controlled elevator should certainly stop at each floor level as an overdamped or critically damped system. The occupants would find it unpleasant if the elevator were to oscillate at each stop.

If $\zeta = 0.5$ the angle ψ is equal to $\psi = \cos^{-1} 0.5 = 60^\circ$. In Fig. 6-13-1a, ω_n is assigned the value of unity and the pole-zero plot of $C(s)$ for a step function input is shown. The corresponding time function can be

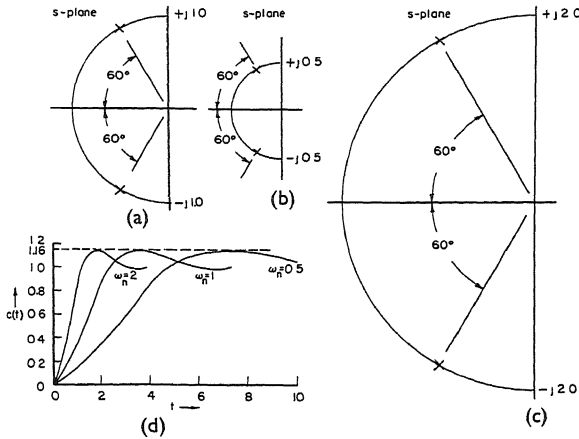


Fig. 6-13-1.

obtained from Fig. 6-11-2 by letting $\omega_n = 1$. The resulting response is shown in Fig. 6-10-1d.

In Fig. 6-13-1b, ω_n is assigned a value of 0.5. The abscissa of Fig. 6-11-2 is $\omega_n t$ which in this case must be multiplied by 2 to yield the corresponding time function of Fig. 6-13-1d. In Fig. 6-13-1c, ω_n is assigned the value of 2. The $\omega_n t$ abscissa of Fig. 6-11-2 must be multiplied by 0.5 to yield the corresponding time function of Fig. 6-13-1d.

The three systems of Fig. 6-13-1d have the same overshoot, the same oscillatory pattern and are said to have the same relative stability.

This discussion can be generalized to closed-loop pole-zero configurations of any complexity. As the locations of the poles and zeros are stretched (or contracted) radially by a factor K in the complex s -plane, the corresponding points on the time function occur at a new time which is the original time multiplied by a factor $1/K$.

It is also informative to examine the response curves for the inputs $u_0(t)$, $u_{-1}(t)$ and $u_{-2}(t)$ as the integral of the previous set of curves. The $e(t)$ curves for $\zeta = 1$ are taken as a specific example. When the input is $u_0(t)$, Fig. 6-10-2 shows that $e(t)$ contains a unit impulse at the origin and is negative thereafter. The integral of this curve is the response to $u_{-1}(t)$ and is contained in Fig. 6-11-2. The integral of the impulse at $t = 0$ yields an $e(t)$ with a unit value at $t = 0$ and the area under the negative portion of the original curve just equals the area of the impulse so that the integral yields an $e(t)$ that approaches zero as t approaches infinity.

When the $e(t)$ curve is integrated again, the result is the response to $u_{-2}(t)$ and is shown in Fig. 6-12-2. The new integral approaches a constant value as t approaches infinity. If this last $e(t)$ curve were integrated

again, the result would go to infinity as t approaches infinity. Thus the error in steady-state for this particular system is infinite when the input is $u_{-3}(t)$. The general subject of steady-state errors is taken up in another article.

6-14. Examples. The following examples illustrate the methods of the previous article.

Example 6-14-1. The system of Fig. 6-7-3 has the following constants

$$\begin{aligned}
 J_{\text{load}} &= 5.18 \times 10^{-3} \text{ in.-lb-sec}^2 \\
 J_{\text{motor}} &= 1.44 \times 10^{-4} \text{ in.-lb-sec}^2 \\
 D_{\text{load}} &= 1.0 \times 10^{-3} \text{ in.-lb-sec} \\
 D_{\text{motor}} &= 0.07 \times 10^{-4} \text{ in.-lb-sec} \\
 K_a &= 0.796 \text{ volt/radian} \\
 K_b &= 20 \text{ volts/volt} \\
 R_a &= 10 \text{ ohms} \\
 K_T &= 9.08 \times 10^{-3} \text{ in.-lb/amp} \\
 K_{av} &= 2.4 \times 10^{-2} \text{ volt-sec/radian} \\
 n &= 100
 \end{aligned} \tag{6-14-1}$$

From these constants, the following values are calculated

$$\begin{aligned}
 J &= J_l + J_m n^2 = 1.445 \text{ in.-lb-sec}^2 \\
 D &= D_l + D_m n^2 + \frac{K_{av} K_T}{R_a} n^2 = 0.289 \text{ in.-lb-sec} \\
 K &= \left(\frac{K_a K_b K_T}{R_a} \right) n = 1.448 \text{ in.-lb/radian} \\
 \omega_n &= \sqrt{K/J} \approx 1 \text{ radian/sec} \\
 \zeta &= \frac{D}{2\sqrt{KJ}} = 0.1
 \end{aligned} \tag{6-14-2}$$

If the system is at rest and the input shaft is suddenly turned through an angle of θ_1 radians, the output and actuating signal can be obtained from Fig. 6-11-2. Since in this example $\omega_n = 1$ the horizontal scale need not be changed. However the vertical scale must be multiplied by θ_1 . The equation for $c(t)$ can be found from Eq. 6-11-11 as

$$c(t) = \theta_1 [1 - 1.005e^{-0.1t} \sin(0.995t + 84.27)] \tag{6-14-3}$$

If the system is at rest and the input shaft is suddenly turned at a constant angular velocity ω_1 the output and the actuating signal can be obtained from Fig. 6-12-1 and Fig. 6-12-2 respectively. Again, because $\omega_n = 1$, the horizontal scale need not be changed. However the vertical scale must be multiplied by $\omega_1/\omega_n = \omega_1$.

Because of the low value of ζ , the system response to any type of input is highly oscillatory. This example points out the fact that the designer must be able to vary at least some of the system parameters in order to improve the response. This discussion is carried on into the next example.

Example 6-14-2. The system of Ex. 6-14-1 is to be redesigned so that the relative stability is improved. It is decided that a value of $\zeta = 0.5$ yields a satisfactory relative stability. However, the redesigned system should have the same steady-

state error when a unit-ramp function is applied to the input. Of the three parameters, J , D , and K , J is generally not easily variable and is assumed fixed but the other two are the variables.

From the generalized equation, the steady-state error with a unit-ramp input for the system of Ex. 6-14-1 is

$$e(t)_{ss} = \frac{2\zeta}{\omega_n} = \frac{2(0.1)}{1} = 0.2 \text{ radian} \quad (6-14-4)$$

The redesigned system is to maintain this same steady-state error but with a $\zeta = 0.5$

$$e(t)_{ss} = \frac{2\zeta}{\omega_n} = \frac{2(0.5)}{\omega_n} = 0.2 \text{ radian} \quad (6-14-5)$$

which yields

$$\omega_n = 5 \text{ radians/sec} \quad (6-14-6)$$

The defining equation for ω_n is solved for K as

$$K = \omega_n^2 J = 36.1 \text{ in.-lb radian} \quad (6-14-7)$$

and D is determined from

$$D = 2\zeta\sqrt{KJ} = 7.23 \text{ in.-lb-sec} \quad (6-14-8)$$

Both the new K and D are 25 times their original values. One factor that goes to make up K is K_b , the gain of the amplifier. Without considering the effects on the motor, etc., caused by increasing the gain, it seems reasonable that the gain of the amplifier can be increased even though a new stage of amplification may be required. However, the increase in D is not a simple matter and is discussed in the next article.

6-15. Damping Increased by Using Feedback from Output Shaft.

As discussed in the previous example, it may be desired to vary D as a design parameter. One possibility is to increase the actual viscous friction by using a device such as a dash pot. When D is made up entirely of viscous friction the power dissipated in the form of heat is

$$\text{Power} = (\text{Torque}) \times (\text{Angular velocity}) = D \left(\frac{dc}{dt} \right)^2 \quad (6-15-1)$$

any increase in D will produce a corresponding power loss and is undesirable.

The equation for D for the system of Fig. 6-7-3 is that of Eq. 6-7-18 repeated as

$$D = \left(D_m + \frac{K_{av} K_T}{R_a} \right) n^2 \quad (6-15-2)$$

This equation suggests that D can be increased by an increase in one of the K 's of the second term. Since K_T affects both D and K , it is desirable to increase K_{av} . This is done by placing a tachometer on the motor shaft and feeding the voltage back as shown in the block-diagram of Fig. 6-15-1

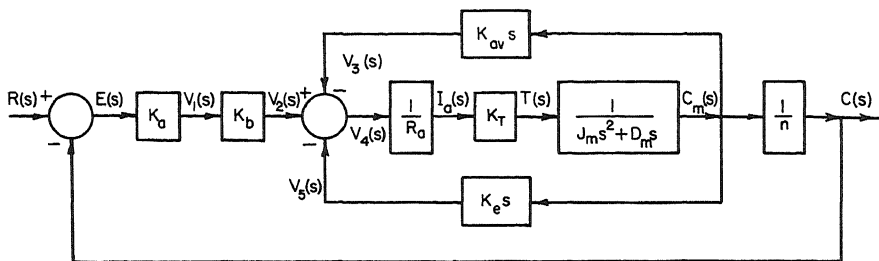


Fig. 6-15-1.

with K_e the constant of proportionality between motor shaft angular velocity and the voltage $v_5(t)$. Since the feedback path of the minor loops are in parallel, the transfer functions can be added

$$(K_{av} + K_e)s \quad (6-15-3)$$

Therefore the K_{av} of the original system is increased to $K_{av} + K_e$ due to the tachometer and D is also increased as

$$D = \left[D_m + \frac{(K_{av} + K_e)K_T}{R_a} \right] n^2 \quad (6-15-4)$$

The advantage of increasing D is that a large value of ζ can be obtained

$$\zeta = \frac{D}{2\sqrt{KJ}} \quad (6-15-5)$$

but the disadvantages is that the steady-state error to a unit-ramp input also increases

$$e(t)_{ss} = \frac{D}{K} \quad (6-15-6)$$

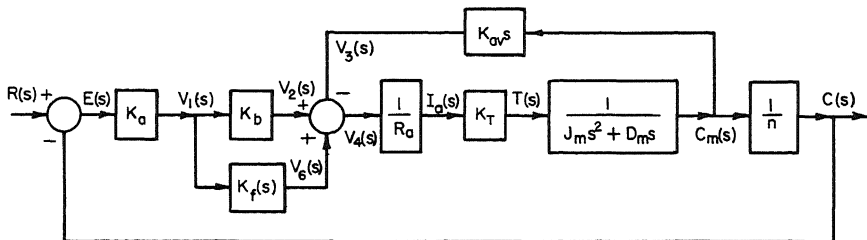


Fig. 6-16-1.

6-16. Error Rate Damping. The system shown in Fig. 6-16-1 is a modification of the system of the previous article. The signal $V_1(s)$ divides along two parallel paths which have a combined transfer function

$$K_b + K_f s \quad (6-16-1)$$

The equations developed in Ex. 6-7-1 hold if K_b is replaced by $K_b + K_f s$. The equations for D and J do not contain K_b and are unaffected by the change; however, K from Eq. 6-7-15 repeated as

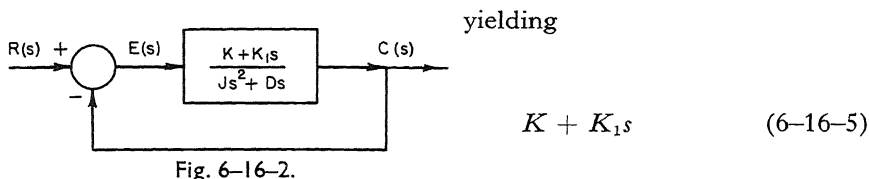
$$K = \frac{K_a K_b K_T}{R_a} n \quad (6-16-2)$$

becomes

$$\frac{K_a(K_b + K_f s)K_T}{R_a} n = \frac{K_a K_b K_T}{R_a} n + \frac{K_a K_f K_T}{R_a} ns \quad (6-16-3)$$

The factor multiplying s is replaced by K_1

$$K_1 = \frac{K_a K_f K_T}{R_a} n \quad (6-16-4)$$



The block diagram of Fig. 6-7-4 can be replaced by one in which K is replaced by $K + K_1 s$ as shown in Fig. 6-16-2.

The voltage $v_6(t)$ can be obtained by differentiating $v_1(t)$. Another rather crude method of obtaining $v_6(t)$ is the output voltage from a tachometer placed on the error shaft. As explained in Chapter 8, this type of voltage in practice is approximated by the output from a differentiating (lead) network.

The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{K + K_1 s}{Js^2 + (D + K_1)s + K} \quad (6-16-6)$$

and the ratio of $E(s)$ to $R(s)$ is

$$\frac{E(s)}{R(s)} = \frac{Js^2 + Ds}{Js^2 + (D + K_1)s + K} \quad (6-16-7)$$

When the input to the system is a unit-ramp function, $E(s)$ becomes

$$E(s) = \frac{Js + D}{s[Js^2 + (D + K_1)s + K]} \quad (6-16-8)$$

and the final-value theorem yields for

$$e(t)_{ss} = \frac{D}{K} \quad (6-16-9)$$

whereas ζ is now given by

$$\zeta = \frac{D + K_1}{2\sqrt{KJ}} \quad (6-16-10)$$

In other words, the designer can now control ζ by varying K_1 without affecting the ratio of D/K .

Example 6-16-1. The requirements of Ex. 6-14-2 are used to design a system with error rate damping. The $e(t)_{ss}$ of Eq. 6-16-9 is to equal 0.2 and the D and K from Ex. 6-14-1 satisfy this requirement without change

$$0.2 = \frac{D}{K} = \frac{0.289}{1.448} \quad (6-16-11)$$

The value of $\zeta = 0.5$ is attained by adding the necessary K_1 to D as

$$\zeta = 0.5 = \frac{D + K_1}{2\sqrt{KJ}} = \frac{0.289 + K_1}{2\sqrt{1.448 \times 1.445}} \quad (6-16-12)$$

from which K_1 is found as

$$K_1 \approx 1.156 \quad (6-16-13)$$

In Ex. 6-14-2, both K and D are increased by a factor of 25 and ω_n by a factor of 5. In this example, K , D , and ω_n are unchanged.

The system with error rate damping is still a relatively simple system and yet the possibility of plotting generalized curves is remote. An examination of Eq. 6-16-10, repeated as,

$$\zeta = \frac{D + K_1}{2\sqrt{KJ}} \quad (6-16-14)$$

indicates the problem. For each value of ζ an entire family of curves results depending on how much of $D + K_1$ is made up of D (or of K_1).

In terms of poles and zeros, the closed-loop transfer function for this system has two poles and one finite zero. For any fixed location of the poles, an entire family of curves result depending on the location of the zero.

6-17. Other Derivative Compensation Schemes. It is also possible to take the derivative of the input-signal $r(t)$ and introduce it into the input to the amplifier with either a positive or negative polarity. Such a procedure does not alter the damping ratio, ζ ; however, it is possible to eliminate the steady-state error to ramp inputs if the positive polarity is used. This analysis is left as Prob. 6-47.

It is theoretically possible to alter the system performance by using the second derivatives of signals. Such procedure would affect the inertia (or mass) terms in the equations. However, these methods are not of sufficient importance to warrant further consideration in this book.

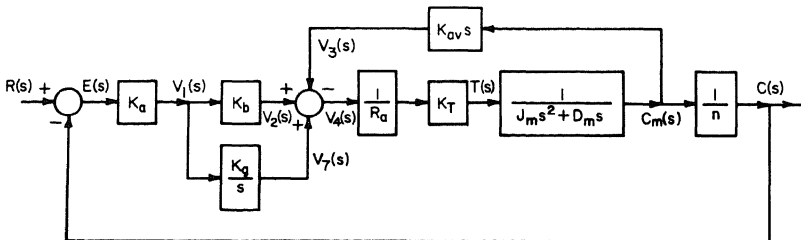


Fig. 6-18-1.

6-18. Integral Compensation. The previous articles outline a few of the methods available to the designer for manipulating system parameters. The method of varying the damping by using error rate compensation is superior to the other methods considered because the designer can increase ζ without increasing D/K the steady-state error to a unit-ramp input. Even though this method does not increase D/K , it also does not decrease this ratio. The system of Fig. 6-18-1 is a modification of the system of the previous articles. The signal $V_1(s)$ divides along two parallel paths which have a combined transfer function

$$K_b + \frac{K_g}{s} \quad (6-18-1)$$

The equations developed in Ex. 6-7-1 hold if K_b is replaced by $K_b + K_g/s$. The equations for D and J do not contain K_b and are unaffected by the change: however K from Eq. 6-7-5 repeated as

$$K = \frac{K_a K_b K_T}{R_a} n \quad (6-18-2)$$

becomes

$$\frac{K_a(K_b + K_g/s)K_T n}{R_a} = \frac{K_a K_b K_T n}{R_a} + \frac{K_a K_g K_T n}{R_a s} \quad (6-18-3)$$

The factor multiplying $1/s$ is replaced by K_2

$$K_2 = \frac{K_a K_g K_T n}{R_a} \quad (6-18-4)$$

yielding

$$K + \frac{K_2}{s} \quad (6-18-5)$$

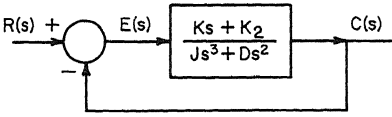


Fig. 6-18-2.

The block diagram of Fig. 6-7-4 can be replaced by one in which K is replaced by $K + K_2/s$ as shown in Fig. 6-18-2.

The voltage $v_7(t)$ is obtained by integrating $v_1(t)$. As explained in Chapter 8, this type of voltage in practice is approximated by the output of an integrating (lag) network.

The closed-loop transfer function is

$$\frac{C(s)}{R(s)} = \frac{Ks + K_2}{Js^3 + Ds^2 + Ks + K_2} \quad (6-18-6)$$

and the ratio of $E(s)$ to $R(s)$ is

$$\frac{E(s)}{R(s)} = \frac{Js^3 + Ds^2}{Js^3 + Ds^2 + Ks + K_2} \quad (6-18-7)$$

When the input to the system is a unit-ramp function, $E(s)$ becomes

$$E(s) = \frac{Js + D}{Js^3 + Ds^2 + Ks + K_2} \quad (6-18-8)$$

and the final-value theorem yields

$$e(t)_{ss} = 0 \quad (6-18-9)$$

The use of this theorem assumes that none of the poles of Eq. 6-18-8 are in the right half s -plane or on the $j\omega$ -axis, that is, it assumes stability of the closed-loop system. The possibility of producing an unstable system is the price paid for reducing the steady-state error to zero with a ramp input. It is shown in Chapter 7 that the closed-loop system is stable if

$$KD > K_2J \quad (6-18-10)$$

This means that if K_2 is made larger than KD/J , the system becomes unstable. Also as discussed in Chapters 7 and 8, even though K_2 is maintained below the limit of absolute stability any increase in K_2 produces a system of lower relative stability.

Error rate damping and integral compensation can be combined in a single system. The development of the system equation is left as Prob. 6-27.

Example 6-18-1. The following parameters are assumed for a system with integral compensation

$$\begin{aligned} J &= 0.4 \text{ in.-lb-sec}^2 \\ D &= 6 \text{ in.-lb-sec} \\ K &= 41.6 \text{ in.-lb/radian} \\ K_2 &= 124.8 \text{ in.-lb-sec/radian} \end{aligned} \quad (6-8-11)$$

Before examining the system as specified, the simpler second-order system with K_2 set equal to zero is first explored in order to compare results. The $C(s)$ for a unit-step input is given by

$$C(s) = \frac{104}{s[s^2 + 15s + 104]} = \frac{104}{s[(s + 7.5)^2 + (6.91)^2]} \quad (6-18-12)$$

Item 19 in the Transform Table is used to find $c(t)$ as

$$c(t) = 1 - 1.477e^{-7.5t} \sin(6.91t + 42.7^\circ) \quad (6-18-13)$$

When K_2 is added, $C(s)$ becomes

$$C(s) = \frac{104s + 312}{s[s^3 + 15s^2 + 104s + 312]} \quad (6-18-14)$$

The denominator can be factored by methods discussed in the next article and $C(s)$ is written as

$$C(s) = \frac{104s + 312}{s[s + 6.36][(s + 4.32)^2 + (5.51)^2]} \quad (6-18-15)$$

Item 25 in the Transform Table yields

$$c(t) = 1 + 1.588e^{-6.36t} + 2.6e^{-4.32t} \sin(5.51t - 94.3^\circ) \quad (6-18-16)$$

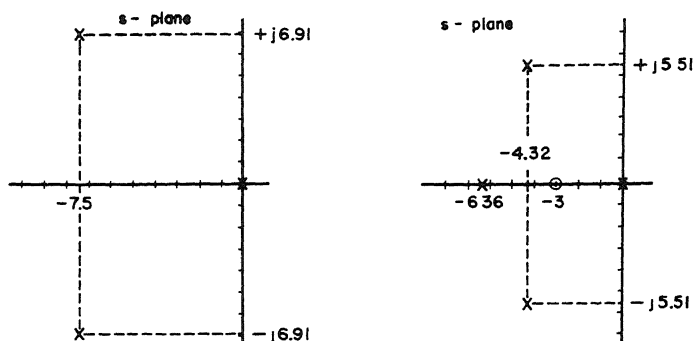


Fig. 6-18-3.

The pole-zero pattern for the system with $K_2 = 0$ is shown in Fig. 6-18-3a and for the system with $K_2 = 124.8$ in Fig. 6-18-31b. The two time functions are shown in Fig. 6-18-4. An examination of these time functions reveals that the system

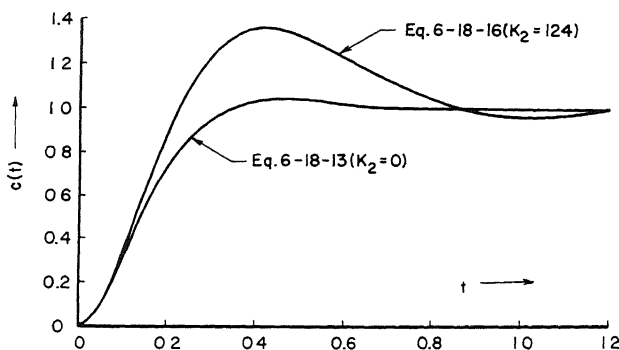


Fig. 6-18-4.

with integral compensation has a higher overshoot, more pronounced oscillations and in general a lower relative stability than the system with $K_2 = 0$. In terms of poles and zeros, the integral compensation adds a pole and a zero to the closed-loop transfer function in addition to the original pair of poles. The zero is inside the pole (closer to the origin) and produces a greater overshoot as discussed in Chapter 10. In addition to this the complex pair of poles is moved closer to the $j\omega$ -axis and the angle from one of the complex poles to the negative real axis has also increased.

As K_2 is increased above the value of 124.8, the complex pair of poles moves closer and closer to the $j\omega$ axis and when $K_2 = KD/J = 624$ the pair of poles lies on the $j\omega$ -axis at $\pm j 10.2$ and system oscillates indefinitely. If K_2 is increased above this value, the poles move into the right half s -plane, the system becomes unstable and the oscillations build up until some component of the system saturates or in some cases is damaged.

6-19. Transient Response of Higher-Order Systems. The characteristic polynomial for most practical systems is of fairly high order. This means that determining the transient response to any kind of an input is a laborious task. One of the difficulties is that the characteristic polynomial must be factored. There are no known methods of factoring polynomials of higher than the 5th degree by formula. However, there

are a number of methods for approximating the zeros of polynomials. Lin's method is featured here because of its basic simplicity. The procedure is demonstrated by an example.

Example 6-19-1. The $C(s)$ for a system with a step-function input is given by

$$C(s) = \frac{32,400}{s[s^5 + 67s^4 + 2,244s^3 + 14,058s^2 + 44,280s + 32,400]} \quad (6-19-1)$$

The denominator is a polynomial of the 5th degree and can be partially factored into two quadratic terms and a linear term. The quadratic formula can be used on the quadratic terms even though the zeros are complex. Lin's method starts by obtaining a trial divisor and after a division process a second trial divisor is obtained and the process is repeated until a desired degree of accuracy is achieved.

In this example, since the degree of the polynomial is odd, the first trial divisor is a linear term obtained from the last two terms of the polynomial as

$$44,280s + 32,400 = 44,280(s + 0.73) \quad (6-19-2)$$

and the division is carried out as

$$\begin{array}{r}
 s^4 + 66.27s^3 + 2,196s^2 + 12,450s + 35,170 \\
 s + 0.73 \overline{) s^5 + 67s^4 + 2,244s^3 + 14,058s^2 + 44,250s + 32,400} \\
 \underline{s^5 + 0.73s^4} \\
 66.27s^4 + 2,244s^3 \\
 \underline{66.27s^4 + 48.5s^3} \\
 2,196s^3 + 14,058s^2 \\
 \underline{2,196s^3 + 1,608s^2} \\
 12,450s^2 + 44,280s \\
 \underline{12,450s^2 + 9,110s} \\
 \text{Second trial divisor: } 35,170s + 32,400 \\
 \underline{35,170s + 25,700} \\
 \text{Remainder: } 6,700
 \end{array} \quad (6-19-3)$$

If the trial divisor were exact, the remainder would be zero instead of 6,700. A rough indication of the accuracy of the $s + 0.73$ term is obtained by comparing the remainder 6,700 with the coefficients 32,400 and 25,700 from which it is obtained. Since the remainder is approximately 20% of these terms, a second trial divisor is obtained from the location indicated

$$35,170s + 32,400 = 35,170(s + 0.92) \quad (6-19-4)$$

The term $s + 0.92$ is used and the division process repeated. Generally the remainder decreases and the trial and error process converges. In the present example after four divisions the polynomial is factored as

$$(s + 1)(s^4 + 66s^3 + 2,178s^2 + 11,880s + 32,400) \quad (6-19-5)$$

If the quartic polynomial has four complex roots, a division by a linear term cannot possibly lead to any useful results. Therefore the last three terms are used to form a quadratic trial divisor and the following division is carried out

$$\begin{array}{r}
 s^2 + 60.5s + 1,830 \\
 s^2 + 5.5s + 15 \overline{) s^4 + 66.0s^2 + 2,178s^2 + 11,880s + 32,400} \\
 \underline{s^4 + 5.5s^3 + 15s^2} \\
 60.5s^3 + 2,163s^2 + 11,880s \\
 \underline{60.5s^3 + 333s^2 + 907s} \\
 \text{Second trial divisor: } 1,830s^2 + 10,973s + 32,400 \\
 \underline{1,830s^2 + 10,073s + 27,400} \\
 \text{Remainder: } 900s + 5,000
 \end{array} \quad (6-19-6)$$

The new trial divisor is now formed as before and the process repeated. After three division processes, the equation is factored as

$$(s + 1)(s^2 + 6s + 18)(s^2 + 60s + 1,800) \quad (6-19-7)$$

The quadratic terms are factored by using the quadratic formula and $C(s)$ put in the form

$$C(s) = \frac{32,400}{s[s + 1][(s + 3)^2 + (3)^2][(s + 30)^2 + (30)^2]} \quad (6-19-8)$$

In some cases, the process of repeated trial divisors will not converge to a solution or in other cases the process converges so slowly other methods of factoring must be used. An alternate method of factoring is explained in Chapter 7. However for the situation in which Lin's method converges rapidly, it is probably the best method. Therefore it is advisable to try Lin's method before attempting other methods.

Even though the Transform Table in the Appendix is relatively extensive, the transform of Eq. 6-19-8 is not included. This example points out the fact that regardless of the inclusiveness of a table there will still be occasions when the designer must return to partial fraction expansion or some variation of it. The time function corresponding to $C(s)$ is ground out as

$$c(t) = 1 - 1.43e^{-t} + 0.432e^{-3t} \sin(3t - 264.5^\circ) + 0.000375e^{-30t} \sin(30t - 173.2^\circ) \quad (6-19-9)$$

The pole-zero pattern for this system is shown in Fig. 6-19-1. By analogy to the second order system each pair of complex conjugate poles can be assigned a damping ratio and a natural radian frequency of oscillation determined geometrically as before. The pole pair at $s = -3 \pm j3$ has

$$\zeta_1 = \cos 45^\circ = 0.707 \quad (6-19-10)$$

$$\omega_{n1} = 3\sqrt{2} = 4.242 \quad (6-19-11)$$

and the pole pair at $s = -30 \pm j30$ has

$$\zeta_2 = \cos 45^\circ = 0.707 \quad (6-19-12)$$

$$\omega_{n2} = 30\sqrt{2} = 42.42 \quad (6-19-13)$$

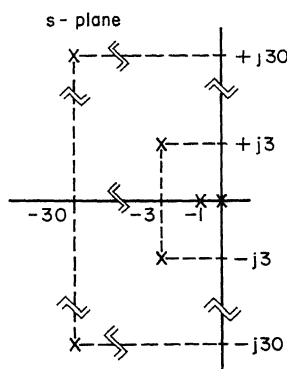


Fig. 6-19-1

This process is used more as a rough analogy to the second-order system than as an exact method of analysis.

This example is reviewed again as an introduction to the concept of dominant poles, a subject that is discussed in more detail in this and other chapters. An examination of Eq. 6-19-9 indicates that the term due to the pole pair at $s = -30 \pm j30$ does not contribute a great deal to the response as the 0.000375 coefficient is small compared with 0.432 and also because this term decays more rapidly than does the other term. In relation to the pole-zero pattern, this pole pair is far removed from the other poles. If the poles at $s = -30 \pm j30$ are dropped from $C(s)$ of Eq. 6-19-8 and the gain of the system reset so that the final value of $c(t)$ is unity, $C(s)$ becomes

$$C(s) = \frac{18}{s[s + 1][(s + 3)^2 + (3)^2]} \quad (6-19-14)$$

and the resulting $c(t)$ is

$$c(t) = 1 - 1.385e^{-t} + 0.391e^{-3t} \sin(3t - 258.7^\circ) \quad (6-19-15)$$

The $c(t)$ of Eq. 6-19-9 and the approximation to this response as given by Eq. 6-19-15 are plotted in Fig. 6-19-2.

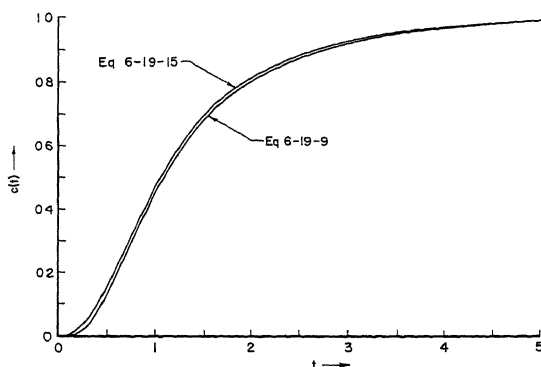


Fig. 6-19-2.

6-20. Static Error Coefficients. Constants that are called static error coefficients are defined for the unity feedback system in the following way. The position constant K_p is defined as

$$K_p = \lim_{s \rightarrow 0} G(s) \quad (6-20-1)$$

the velocity constant K_v as

$$K_v = \lim_{s \rightarrow 0} sG(s) \quad (6-20-2)$$

and the acceleration constant K_a as

$$K_a = \lim_{s \rightarrow 0} s^2 G(s) \quad (6-20-3)$$

The transform for the error signal is

$$E(s) = \frac{R(s)}{1 + G(s)} \quad (6-20-4)$$

When $r(t) = u_{-1}(t)$, $E(s)$ becomes

$$E(s) = \frac{1}{s[1 + G(s)]} \quad (6-20-5)$$

and the steady-state error can be found using the final-value theorem as

$$e(t)_{ss} = \lim_{s \rightarrow 0} \left[\frac{1}{1 + G(s)} \right] = \frac{1}{1 + K_p} \quad (6-20-6)$$

When $r(t) = u_{-2}(t)$, $E(s)$ becomes

$$E(s) = \frac{1}{s^2[1 + G(s)]} \quad (6-20-7)$$

and the steady-state error is

$$e(t)_{ss} = \lim_{s \rightarrow 0} \left[\frac{1}{s + sG(s)} \right] = \frac{1}{K_v} \quad (6-20-8)$$

When $r(t) = u_{-3}(t)$, $E(s)$ becomes

$$E(s) = \frac{1}{s^3[1 + G(s)]} \quad (6-20-9)$$

and the steady-state error is

$$e(t)_{ss} = \lim_{s \rightarrow 0} \left[\frac{1}{s^2 + s^2G(s)} \right] = \frac{1}{K_a} \quad (6-20-10)$$

The valid use of the final-value theorem, in each case depends upon the system's being stable in order that each limit exist. These constants are discussed after another set of definitions is introduced in the next article.

Another set of constants called the dynamic error coefficients are explained in Chapter 10.

6-21. Type Numbers for Control Systems. The forward loop of a unity feedback system has a transfer function of the following form

$$G(s) = \frac{K(s + z_1) \dots (s + z_k)}{s^N(s + p_1) \dots (s + p_j)} \quad (6-21-1)$$

TABLE 6-21-1

Value of N	System type number
$N = 0$	Type 0
$N = 1$	Type 1
$N = 2$	Type 2
$N = 3$	Type 3

The explicit numbers of z and p terms are not pertinent to the present discussion which is concerned only with N . The type numbers correspond with N as given in Table 6-21-1.

With regard to nomenclature, the table can be carried farther. However, as explained in Chapter 7, it is difficult to achieve a Type 3 or higher system that is stable.

TABLE 6-21-2

	Type 0	Type 1	Type 2
K_p	K_p (finite)	∞	∞
K_v	0	K_v (finite)	∞
K_a	0	0	K_a (finite)

The static error coefficients for systems of various type numbers are shown in Table 6-21-2. For a given type number, each error coefficient is either infinite, finite (the value depends on the actual system involved), or zero.

The static error coefficients of Table 6-21-2 are combined with Eq. 6-20-6, 6-20-7 and 6-20-8 which relate the steady-state error and the error coefficients. The result is shown in Table 6-21-3.

TABLE 6-21-3

Input function	Steady-state error		
	Type 0	Type 1	Type 2
$u_{-1}(t)$	$\frac{1}{1 + K_p}$	0	0
$u_{-2}(t)$	∞	$\frac{1}{K_v}$	0
$u_{-3}(t)$	∞	∞	$\frac{1}{K_a}$

Table 6-21-3 is examined by rows. The important point in this discussion is that $G(s)$ has an s^N factor in the denominator which corresponds to N integrations in the time domain.

The first row in this table is for the input $r(t) = u_{-1}(t)$. A Type 0 system does not integrate the error signal. Therefore, for the output to

maintain a constant value, $e(t)_{ss}$ must also have a constant value. A type 1 system integrates the error signal. Therefore, the output can maintain a constant value if the integral of $e(t)$ is equal to a constant. For the integral of $e(t)$ to be a constant, $e(t)_{ss}$ must equal zero. The same type of reasoning shows that $e(t)_{ss}$ is also zero for a Type 2 system.

The second row of Table 6-21-3 is for $r(t) = u_{-2}(t)$. For the output of a Type 0 system to be increasing at a constant rate, the $e(t)$ signal must also be increasing at a constant rate and $e(t)_{ss}$ is therefore equal to infinity. A Type 1 system integrates $e(t)$ and the integral of $e(t)$ in steady state can increase at a uniform rate as long as $e(t)_{ss}$ is equal to a constant. A Type 2 system integrates $e(t)$ two times and the double integral of $e(t)$ can increase in steady state at a uniform rate with $e(t)_{ss}$ equal to zero.

The third row of Table 6-21-3 is for $r(t) = u_{-3}(t)$. For the output of a Type 0 system to be increasing with a constant acceleration, the $e(t)$ signal must also be increasing with a constant acceleration and $e(t)_{ss}$ is therefore equal to infinity. The $e(t)$ signal for a Type 1 system must increase at a constant rate for the output to have a constant acceleration. Therefore $e(t)_{ss}$ is equal to infinity. A Type 2 system integrates $e(t)$ two times; and if $e(t)_{ss}$ is equal to a constant, the double integral of this in steady state yields the desired output.

This table can be extended to systems with higher type numbers and to inputs such as $u_{-4}(t)$, etc. The finite steady-state error term appears on the diagonal with zeros above the diagonal and infinities below.

6-22. Discussion of Transient Analysis. If a specific system is given and the complete response to a given input is desired, the transient method of analysis may be laborious, but it is straight forward. The difficulty with the transient method arises when the problem is design rather than analysis. This difficulty in turn is a result of the devious relationship between coefficients of a high degree polynomial and the location of its zeros. An example may help explain this situation. A designer is working with a control system which has an unsatisfactory response to a unit-step function. The designer faces the problem of changing some of the system parameters in order to improve the response. Without augmenting the transient approach by other methods, about the best he can do is to change one or more of the parameters and recalculate the response, repeating the process on a trial-and-error basis until the desired response is obtained.

It was for these reasons that sinusoidal steady-state methods were developed in which the frequency is varied. Chapters 7, 8, and 9 examine these methods in terms of the Nyquist and Bode plots and how these plots can be used in control system design. These frequency response methods themselves suffer limitations and other methods of design are explained in later chapters. An introduction to some of the frequency response concepts is given in the following articles.

6-23. Sinusoidal Steady-State Response. If a sinusoidal input is applied to a system, the steady-state response can be found by methods similar to those explained in Chapter 3. In these methods, the input and the output are replaced by phasors and s is replaced by $j\omega$ throughout the remainder of the expression.

This procedure is followed for the two pole system with the transfer function given by

$$\frac{C(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (6-23-1)$$

in which s is replaced by $j\omega$

$$\frac{C(j\omega)}{R(j\omega)} = \frac{\omega_n^2}{\omega_n^2 - \omega^2 + j2\zeta\omega_n\omega} \quad (6-23-2)$$

In order to put this equation in generalized form, the numerator and denominator are divided by ω_n^2 .

$$\frac{C(j\omega)}{R(j\omega)} = \frac{1}{1 - \left(\frac{\omega}{\omega_n}\right)^2 + j2\zeta\left(\frac{\omega}{\omega_n}\right)} \quad (6-23-3)$$

The ratio of ω to ω_n is denoted by γ as

$$\gamma = \frac{\omega}{\omega_n} \quad (6-23-4)$$

and the equation can be written as

$$\frac{C(j\gamma)}{R(j\gamma)} = \frac{1}{1 - \gamma^2 + j2\zeta\gamma} = M \angle \alpha \quad (6-23-5)$$

In the literature, there is some confusion in nomenclature concerning the ratio of $C(j\gamma)$ over $R(j\gamma)$ but for present purposes the magnitude of Eq. 6-23-5 is labeled M and the angle α .

Curves of M for various values of ζ are shown in Fig. 6-23-1. When ζ is below 0.707, the M curve increases to a maximum and then decreases. The maximum value of M is denoted M_m and the radian frequency at which maximum occurs is denoted ω_m .

If the magnitude of Eq. 6-23-5 is differentiated with respect to γ and set equal to zero, ω_m is obtained as

$$\omega_m = \omega_n \sqrt{1 - 2\zeta^2} \quad (6-23-6)$$

and upon substitution of ω_m in the equation for M the equation for M_m becomes

$$M_m = \frac{1}{2\zeta \sqrt{1 - \zeta^2}} \quad (6-23-7)$$

This equation is plotted in Fig. 6-23-2.

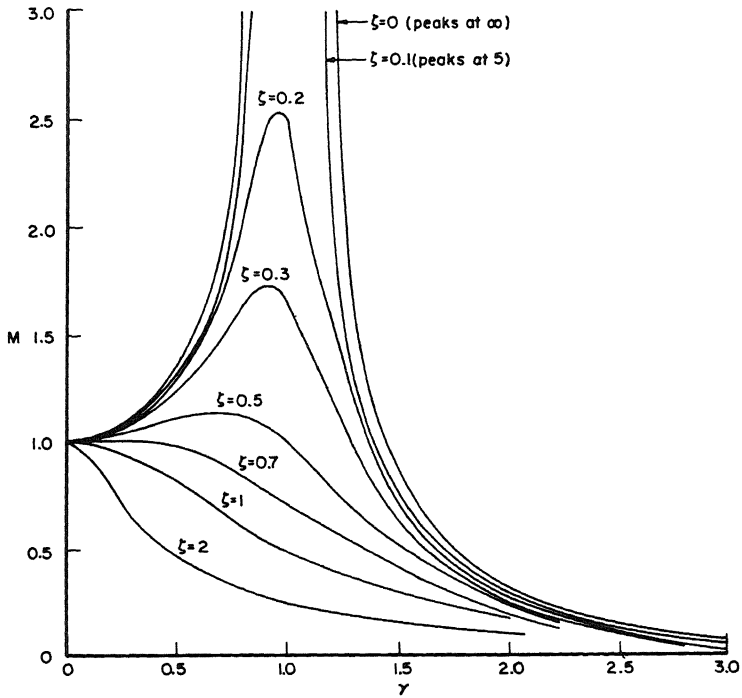


Fig. 6-23-1.

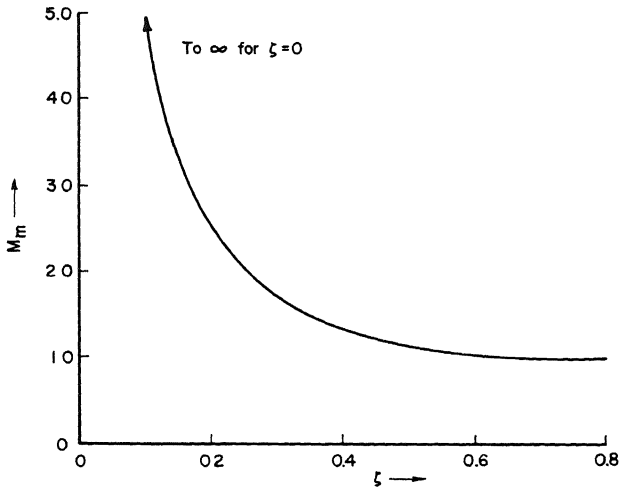


Fig. 6-23-2.

For more complicated systems, it is virtually impossible to plot generalized curves for the same reasons as given in the discussion of transient response.

Example 6-23-1. The control system of Ex. 6-14-1 is driven by the input function $r(t)$ until steady-state conditions are reached.

$$r(t) = 2.5 \sin(0.5t - 55^\circ) - 1.25 \cos(t + 60^\circ) \quad (6-23-8)$$

It is desired to find the equation for the output. The system constants are repeated as

$$D = 0.289, K = 1.448, \omega_n = 1 \text{ and } \zeta = 0.1$$

Table 6-23-1 is set up

TABLE 6-23-1

ω	γ	$\frac{1}{1 - \gamma^2 + j2\zeta\gamma}$
0.5	0.5	$\frac{1}{0.75 + j0.1} = 1.33 \angle -7.6^\circ$
1	1	$\frac{1}{j0.2} = 5 \angle -90^\circ$

The output response is

$$c(t) = 1.33 \times 2.5 \sin(0.5t - 55^\circ - 7.6^\circ) - 1.25 \times 5 \cos(t + 60^\circ - 90^\circ)$$

or

$$c(t) = 3.32 \sin(0.5t - 62.6^\circ) - 6.25 \cos(t - 30^\circ) \quad (6-23-9)$$

Example 6-23-2. A control system has an unknown J and D but tests are made on the controller and the constant K is found to be 2.5. A sinusoidal input of varying frequency is applied and it is found that M_m is 1.5 and ω_m is 1.25 cps. From Fig. 6-23-2, an M_m of 1.5 corresponds to a $\zeta \doteq 0.35$. Eq. 6-23-6 can be solved for ω_n as $\omega_n = 1.44$. J and D are found as $J = K/\omega_n^2 = 1.2$ and $D = 2\zeta\sqrt{KJ} = 1.21$.

6-24. Transient Response Determined by Poles and Zeros.

Both the transient and frequency response methods stem from the same closed-loop transfer function.

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + G(s)} \quad (6-24-1)$$

This transfer function is completely specified in terms of its poles and zeros except for a gain factor. If the control system is of a Type 1 or higher type, the gain is also determined. For example, the following pole-zero pattern is assumed:

$$\frac{C(s)}{R(s)} = \frac{K(s + z_1)}{(s + p_1)(s + p_2)(s + p_3)} \quad (6-24-2)$$

If the input is a unit-step function, $R(s) = 1/s$ and Eq. 6-24-2 can be solved for $C(s)$. For a Type 1 system, it is known that the steady-state

error is zero when a unit-step function is the input. Therefore $c(t)$ in steady state must equal unity which yields

$$c(t)_{ss} = \lim_{s \rightarrow 0} sC(s) = \frac{Kz_1}{p_1 p_2 p_3} = 1 \quad (6-24-3)$$

Upon solving for K , the equation for $C(s)$ becomes

$$C(s) = \frac{p_1 p_2 p_3 (s + z_1)}{z_1 s (s + p_1)(s + p_2)(s + p_3)} \quad (6-24-4)$$

Another factor $s + p_4$ added to the denominator requires a p_4 in the numerator whereas a factor $s + z_2$ added to the numerator requires a z_2 in the denominator, etc.

The partial fraction expansion of $C(s)$ is

$$C(s) = \frac{1}{s} + \frac{K_{p_1}}{s + p_1} + \frac{K_{p_2}}{s + p_2} + \frac{K_{p_3}}{s + p_3} \quad (6-24-5)$$

where the K 's are the residues. The inverse transform of $C(s)$ yields $c(t)$.

$$c(t) = 1 + K_{p_1} \varepsilon^{-p_1 t} + K_{p_2} \varepsilon^{-p_2 t} + K_{p_3} \varepsilon^{-p_3 t} \quad (6-24-6)$$

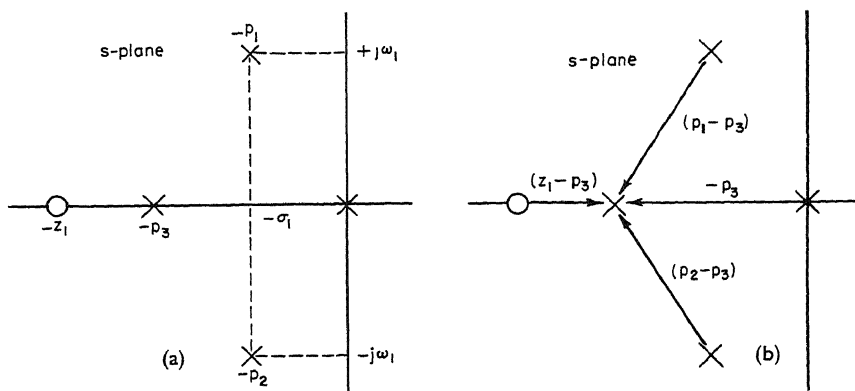


Fig. 6-24-1.

One possible pole-zero configuration is shown in Fig. 6-24-1a. Since p_1 and p_2 are assumed to be complex conjugates $\sigma_1 \pm j\omega_1$, the equation for $c(t)$ can be written as

$$c(t) = 1 + A \varepsilon^{-\sigma_1 t} \sin(\omega_1 t + \psi) + K_{p_3} \varepsilon^{-p_3 t} \quad (6-24-7)$$

The effects of the pole locations on this response can be seen in a qualitative way. The smaller σ_1 , the closer the complex poles are to the $j\omega$ -axis and the longer it takes the damped sinusoidal component to die away. The larger ω_1 , the farther the complex poles are from the real axis and the higher the frequency of oscillation of this component. The farther $-p_3$ is to the left, the more rapidly the term $K_{p_3} \varepsilon^{-p_3 t}$ dies away.

The effects of the zero location in Eq. 6-24-7 are not evident; however, each residue is determined in part by the location of the zero. For example, K_{p_1} is given by

$$K_{-p_1} = \frac{p_1 p_2 p_3 (z_1 - p_3)}{z_1 (-p_3) (p_1 - p_3) (p_2 - p_3)} \quad (6-24-8)$$

The term $z_1 - p_3$ can be visualized as a phasor from $-z_1$ to $-p_3$, as indicated in Fig. 6-24-1b. Likewise the term $p_1 - p_3$ is a phasor from $-p_1$ to $-p_3$. In general, each residue is composed of a group of phasor products in the numerator divided by a group of phasor products in the denominator. For this specific example, the phasor from $-z_1$ to $-p_3$ is in the numerator. As the zero is moved closer to $-p_3$, the K_{-p_1} term becomes smaller. If the zero and pole coincide, the pole is said to be canceled by the zero. Complete cancellation is not necessary for the pole to have little effect because if the zero is near the pole, the residue is small.

Another situation exists in which a pole (or poles) has little effect on the system response. This situation occurs when the pole is far removed from the other poles and zeros. The system response to a step function is used as an example:

$$C(s) = \frac{10}{s(s^2 + s + 1)(s + 10)} \quad (6-24-9)$$

The complex pair of poles is at $-0.5 \pm j0.866$. The inverse transform is

$$c(t) = 1 + 1.21e^{-0.5t} \sin(0.866t - 125.2^\circ) - 0.011e^{-10t} \quad (6-24-10)$$

An examination shows that the pole at -10 does not contribute a great deal to the response as the residue -0.011 is small when compared with 1. If the pole at -10 is dropped, the $C(s)$ becomes

$$C(s) = \frac{1}{s(s^2 + s + 1)} \quad (6-24-11)$$

and the resulting $c(t)$ is

$$c(t) = 1 + 1.16e^{-0.5t} \sin(0.866t - 120^\circ) \quad (6-24-12)$$

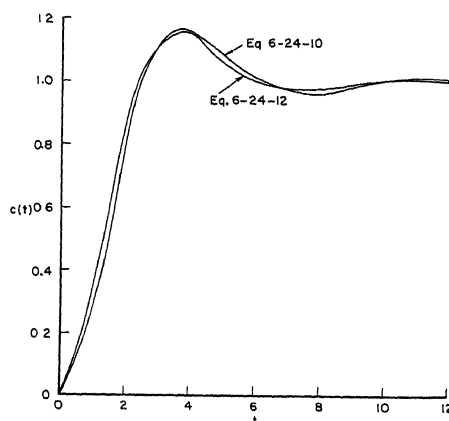


Fig. 6-24-2.

The $c(t)$ of Eq. 6-24-10 and the approximation to this response as given by Eq. 6-24-12 are plotted in Fig. 6-24-2.

The pole-zero pattern and the geometrical interpretation of evaluating the residues are helpful in visualizing the reasons the pole at -10 has a small residue. The residue in the pole at the origin is unity, and therefore the three phasors shown in Fig. 6-24-3a must have a product equal to 10. The phasors in Fig. 6-24-3b determine the residue in the pole at -10 . In

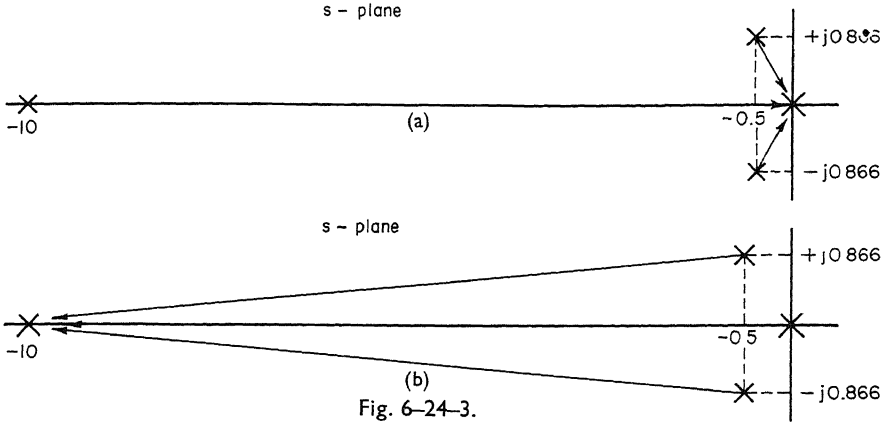


Fig. 6-24-3.

Fig. 6-24-3a the phasors from the complex poles have a magnitude of 1 whereas in Fig. 6-24-3b they have a magnitude of approximately 9.5. Therefore, the product of the three phasors in Fig. 6-24-3b is approximately equal to $(10)(9.5)^2$. If the pole at -10 were farther removed, or if there were additional poles in the vicinity of the unit circle, the residue would be even smaller.

This article introduces the effect of the location of poles and zeros on transient response in a qualitative way. In Chapter 10, this same subject is discussed again in more quantitative terms.

6-25. Sinusoidal Steady-State Response Determined by Poles and Zeros. Not only do the poles and zeros of a transfer function completely determine the transient response, but they also determine the sinusoidal steady-state response. The system described by Eq. 6-24-4 is used again in this discussion.

$$\frac{C(s)}{R(s)} = \frac{p_1 p_2 p_3 (s + z_1)}{z_1 (s + p_1) (s + p_2) (s + p_3)} \quad (6-25-1)$$

The input and output are replaced by phasors and s is replaced by $j\omega$ throughout the remainder of the expression

$$\frac{C(j\omega)}{R(j\omega)} = \frac{p_1 p_2 p_3 (j\omega + z_1)}{z_1 (j\omega + p_1) (j\omega + p_2) (j\omega + p_3)} \quad (6-25-2)$$

The quantity $j\omega$ locates one point on the $j\omega$ -axis in the s -plane and each of the terms $j\omega + p_1$, $j\omega + z_1$, etc. is a phasor from a pole (or a zero) to this point on the $j\omega$ -axis. One such $j\omega$ point and the resulting phasors are shown in Fig. 6-25-1. when ω is equal to zero, $C(j\omega)/R(j\omega)$ is equal

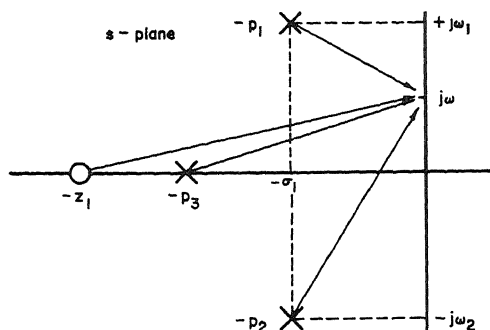


Fig. 6-25-1.

to unity; and as ω varies from zero to infinity, the variations of $C(j\omega)/R(j\omega)$ can be visualized by observing the changes that take place in these phasors.

If the complex pair of poles $-p_1$ and $-p_2$ is close to the $j\omega$ -axis, the frequency response curve is highly peaked. This can be made clear by visualizing $-p_1$ and $-p_2$ close to the $j\omega$ -axis and observing the phasor from $-p_1$ to $j\omega$ as $j\omega$ takes on values in the neighborhood of $j\omega_1$. The phasors from the other poles and zeros do not change very rapidly in this neighborhood, but the phasor from $-p_1$ decreases rapidly and then increases as $j\omega$ moves past $j\omega_1$. Since this phasor is in the denominator, the magnitude of the frequency response increases and then decreases. In general, the peak of this curve does not occur at ω_1 but near ω_1 .

If $-z_1$ is near $-p_3$, the phasors $j\omega + z_1$ and $j\omega + p_3$ are almost identical. Since one phasor is in the numerator and the other in the denominator, it can be seen that a pole and zero close together have little effect on the frequency response.

In the previous article, it is pointed out that a complex pair of poles near the $j\omega$ -axis produces a highly oscillatory transient response and in this article it is shown that this pair of poles produces a highly peaked frequency response. Stated in reverse, if the frequency response is not highly peaked, the pole-zero pattern does not have a pair of complex poles near the $j\omega$ -axis and in turn the transient response will not be highly oscillatory.

For many years a "rule of thumb" has been used to the effect that if M_m (the maximum value of the frequency response curve) is between 1.2 and 1.4, the transient response of the system will be satisfactory. The above discussion qualitatively justifies this "rule of thumb".

Chapter 10 discusses the relationship between transient and frequency response in a more quantitative way.

If the pole-zero pattern is stretched (or contracted) radially by a factor K in the complex s -plane, the corresponding frequency response curve takes on the same set of values but at an ω which is K times the original ω . For example, if the original frequency response curve is obtained from

$$\frac{C(j\omega)}{R(j\omega)} = \frac{p_1 p_2 p_3 (j\omega + z_1)}{(z_1)(j\omega + p_1)(j\omega + p_2)(j\omega + p_3)} \quad (6-25-3)$$

the new curve results from

$$\frac{C(j\omega)}{R(j\omega)} = \frac{(Kp_1)(Kp_2)(Kp_3)(j\omega + Kz_1)}{(Kz_1)(j\omega + Kp_1)(j\omega + Kp_2)(j\omega + Kp_3)} \quad (6-25-4)$$

The original curve has a certain magnitude M for a given ω . The new curve has the same magnitude M at an angular frequency $K\omega$.

This result should be contrasted with the corresponding effect on the transient response. If the pole-zero pattern is stretched by a factor K , the corresponding points on the time function occur at a new time which is the original time multiplied by a factor $1/K$. Therefore, the more spread out the frequency response curve, the more contracted is the corresponding transient response curve and vice versa.

6-26. Transient Response with Initial Conditions. Transfer functions are derived under the assumption that the system is initially at rest. The initial conditions can be added to control systems by use of the impulse family of functions in a manner similar to that discussed in Arts. 3-10 and 3-11. However for this discussion, the more fundamental approach of returning to the differential equation is used. The remainder of the discussion is carried out through examples.

Example 6-26-1. The system of Ex. 6-14-1 used again with $J = 1.445$, $D = 0.289$ and $K = 1.448$. The input is driven at a constant angular velocity of 2.5 radians/sec. until the system reaches steady state and at $t = 0$ the input is arbitrarily stopped at $r(t) = 0$. The transfer function that describes the closed-loop system is

$$\frac{C(s)}{R(s)} = \frac{K}{Js^2 + Ds + K} \quad (6-26-1)$$

This cannot be used as such since the system contains stored energy at $t = 0$; however, the transformed equation can be written as

$$Js^2C(s) + DsC(s) + KC(s) = KR(s) \quad (6-26-2)$$

and the differential equation described by this is

$$J\frac{d^2c(t)}{dt^2} + D\frac{dc(t)}{dt} + Kc(t) = Kr(t) \quad (6-26-3)$$

In this example $r(t)$ is equal to zero by choice. Before $t = 0$, the system has a steady-state error

$$e(t)_{ss} = 2.5 \frac{D}{K} = 0.5 \quad (6-26-4)$$

From $e(t) = r(t) - c(t)$, it can be seen that $c(0) = -0.5$. At $t = 0$, $dc(0)/dt = 2.5$. The transform of the equation with the initial conditions is

$$J[s^2C(s) + s(0.5) - 2.5] + D[sC(s) + 0.5] + KC(s) = 0 \quad (6-26-5)$$

This can be solved for $C(s)$ as

$$C(s) = \frac{-0.5s + 2.4}{s^2 + 0.2s + 1} = \frac{-0.5s + 2.4}{(s + 0.1)^2 + (0.995)^2} \quad (6-26-6)$$

The inverse transform is

$$c(t) = 2.51e^{-0.1t} \sin(0.995t - 11.5^\circ) \quad (6-26-7)$$

Example 6-26-2. The system of Ex. 6-18-1 is used again with $J = 0.4$, $D = 6$, $K = 41.6$ and $K_2 = 124.8$. The input is driven at a constant angular velocity of 2.5 radians/sec until the system reaches steady state and at $t = 0$ the input is arbitrarily stopped at $r(t) = 0$. The transfer function that describes the closed-loop system is obtained from Eq. 6-18-14 as

$$\frac{C(s)}{R(s)} = \frac{104s + 312}{s^3 + 15s^2 + 104s + 312} \quad (6-26-8)$$

The transform of the equation can be written as

$$s^3C(s) + 15s^2C(s) + 104sC(s) + 312C(s) = 104sR(s) + 312R(s) \quad (6-26-9)$$

and the corresponding differential equation is

$$\frac{d^3c(t)}{dt^3} + 15\frac{d^2c(t)}{dt^2} + 104\frac{dc(t)}{dt} + 312c(t) = 104\frac{dr(t)}{dt} + 312r(t) \quad (6-26-10)$$

At $t = 0$, the input is stopped so that $r(0) = 0$ and after $t = 0$, the derivative of $r(t)$ is also zero. Eq. 6-26-10 becomes

$$\frac{d^3c(t)}{dt^3} + 15\frac{d^2c(t)}{dt^2} + 104\frac{dc(t)}{dt} + 312c(t) = 0 \quad (6-26-11)$$

Because of the integral compensation, the output and input are in alignment before $t = 0$ and the output is turning at the same angular velocity as the input as given by $dc(0)/dt = 2.5$. Also before $t = 0$ the second and third derivative of $c(t)$ with respect to time are equal to zero. Since Eq. 6-26-11 is true for all positive values of time, it is also true at $t = 0$. If any of the derivatives of $c(t)$ with respect to time change magnitudes at $t = 0$, the next higher-order derivative must contain an impulse. Since Eq. 6-26-11 does not contain an impulse at $t = 0$ the only derivative that can change is $d^3c(t)/dt^3$. Therefore at $t = 0$

$$c(0) = 0, \quad \frac{dc(0)}{dt} = 2.5 \quad \text{and} \quad \frac{d^2c(0)}{dt^2} = 0 \quad (6-26-12)$$

When these are substituted into Eq. 6-26-11 it becomes

$$\frac{d^3c(0)}{dt^3} + 0 + (104)(2.5) + 0 = 0 \quad (6-26-13)$$

which can be solved as

$$\frac{d^3c(0)}{dt^3} = -(2.5)(104) \quad (6-26-14)$$

The transform of Eq. 6-26-11 is

$$[s^3C(s) - 0 - 2.5s + 0] + 15[s^2C(s) - 0 - 2.5] + 104[sC(s) - 0] + 312C(s) = 0 \quad (6-26-15)$$

from which $C(s)$ is obtained as

$$C(s) = \frac{2.5s + 37.5}{s^3 + 15s^2 + 104s + 312} \quad (6-26-16)$$

The denominator is factored as

$$C(s) = \frac{2.5s + 37.5}{(s + 6.36)[(s + 4.32)^2 + (5.5)^2]} \quad (6-26-17)$$

and $c(t)$ is found as

$$c(t) = 0.935e^{-4.32t} \sin(5.5t - 42.4^\circ) + 0.632e^{-6.36t} \quad (6-26-18)$$

These two examples demonstrate the manner in which initial conditions can be handled by simply returning to the differential equations and by incorporating the initial conditions in the transformed equation. An inspection of these examples shows that the initial condition terms all appear in the numerator of $C(s)$ and that the denominator is unaffected. Therefore, these initial conditions can in no way affect the absolute stability of a system.

Since a given system can be driven before $t = 0$ so as to yield an infinite number of different sets of initial conditions, the value of this effort must be viewed with skepticism. Therefore, from the standpoint of design, it is much better to interpret the specifications in terms of a system initially at rest and to avoid the problem of initial conditions entirely. If in a special case, the response to a certain set of initial conditions must be found, the procedure is straightforward.

PROBLEMS

6-1. Replace the three summing points in tandem in Fig. 6-P-1 by a single summing point. Write the necessary equation to show the equivalence of the two systems.

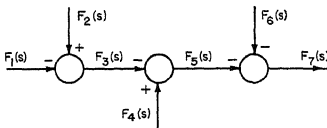


Fig. 6-P-1.

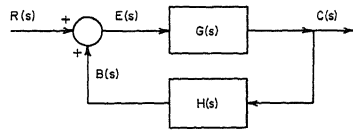


Fig. 6-P-2.

6-2. Given the system of Fig. 6-P-2, write the necessary equations and derive the transfer functions

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 - H(s)G(s)} \quad \text{and} \quad \frac{E(s)}{R(s)} = \frac{1}{1 - H(s)G(s)}$$

6-3. Write the necessary equations to demonstrate that the systems of Fig. 6-5-8a and b yield the same results.

6-4. Repeat Prob. 6-3 except use Fig. 6-5-9.

6-5. Repeat Prob. 6-3 except use Fig. 6-5-10.

6-6. Ex. 6-6-1 reduces the block diagram of Fig. 6-6-1a to a single block by using block diagram transformations. Verify the results obtained in Fig. 6-6-1c by writing a sufficient number of equations relating the variables in Fig. 6-6-1a, and

eliminate all the variables except $C(s)$ and $R(s)$. That is, the result is to be verified by algebraic methods that do not use any of the transformations already derived.

6-7. Repeat Prob. 6-6 except use the systems of Ex. 6-6-2.

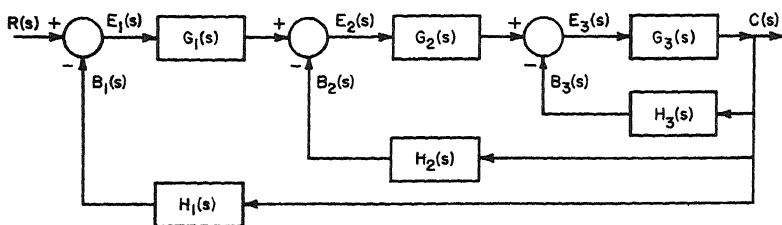


Fig. 6-P-8.

6-8. Given the system of Fig. 6-P-8, reduce this system to a single block diagram transformations.

6-9. Repeat Prob. 6-8. by using algebraic elimination in a manner similar to that suggested in Prob. 6-6.

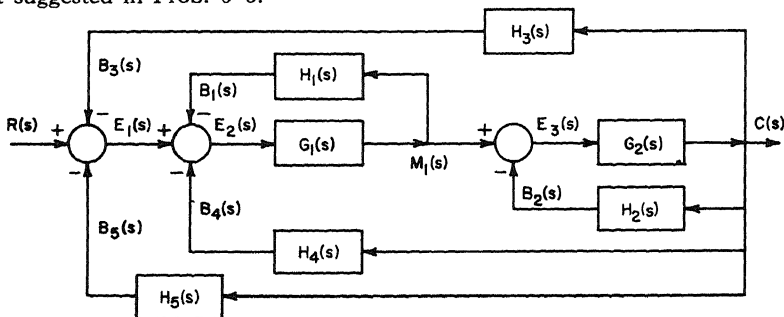


Fig. 6-P-10

6-10. Given the system of Fig. 6-P-10. Reduce this system to a single block by using block diagram transformations.

6-11. Repeat Prob. 6-10 by using algebraic elimination in a manner similar to that suggested in Prob. 6-6.

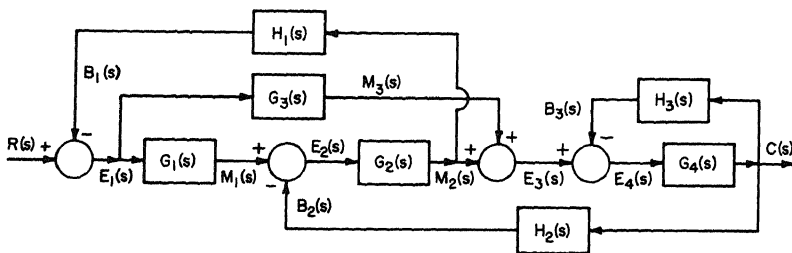


Fig. 6-P-12.

6-12. Given the system of Fig. 6-P-12, reduce this system to a single block by using block diagram transformations.

6-13. Repeat Prob. 6-12 by using algebraic elimination in a manner similar to

6-14. In Ex. 6-6-2 the denominator is made up of 1 plus the sum of the transfer functions around the various loops; however, in Ex. 6-6-1 an extra term is present. Discuss the situation in general as to when this extra term (or terms) will be present and when it will not be present.

6-15. The closed-loop system of Fig. 6-8-1 has the following constants.

$$\begin{aligned} J &= 10 \times 10^{-6} \text{ slug-ft}^2 \\ D &= 2.0 \times 10^{-4} \text{ ft-lb/(radians/sec)} \\ K &= 3.5 \times 10^{-3} \text{ ft-lb/radian} \end{aligned}$$

The input is given a value of 10 for 0.01 sec and then returned to zero. For purposes of analysis, this input is approximated by an impulse which has an area equal to the area under the square pulse. Find $c(t)$ and $e(t)$ and draw rough sketches.

6-16. The system of Prob. 6-15 is initially at rest and the input is suddenly turned through π radians at $t = 0$. Find $c(t)$ and $e(t)$ and draw rough sketches of both functions.

6-17. The system of Prob. 6-15 is initially at rest and the input is suddenly given a constant angular velocity of 3 radians/sec. Find $c(t)$ and $e(t)$ and draw rough sketches of both functions.

6-18. The closed-loop system of Fig. 6-8-1 has $J = 1.445 \text{ in.} \cdot \text{lb-sec}^2$ but D and K are variables. Find D and K so that the system has an $\omega_n = 2.0$ and an overshoot to a unit-step function equal to 30% . Find $c(t)$ and $e(t)$ when $r(t) = u_{-1}(t)$ and draw rough sketches of both functions.

6-19. The closed-loop system of Fig. 6-8-1 has $J = 1.445 \text{ in.} \cdot \text{lb-sec}^2$ but D and K are variables. It is desired to have the system operate with a relative stability as determined by $\zeta = 0.5$. The system is initially at rest and is driven with a constant angular velocity of 3 radians/sec. The error in steady state is to be 0.5 radian. Find $c(t)$ and $e(t)$ and draw rough sketches of both functions.

6-20. The control system of Ex. 6-14-1 is redesigned in Ex. 6-14-2. This is done by increasing K and D . K is increased by an increase in K_b . Find the new value of K_b required. D is increased by using feedback from the output shaft as shown in Fig. 6-15-1. Find the value of K_e required.

6-21. A control system is designed as discussed in Prob. 6-20. However the signal $V_5(s)$ is inadvertently connected so that it passes through the summing point with a plus sign instead of the minus sign shown in Fig. 6-15-1. Calculate the step function response of the resulting system.

6-22. The control system of Prob. 6-21. is left so that $V_5(s)$ passes through the summing point with a plus sign; however, K_e is set so that $D = 0$. Find the value of K_e to do this. The resulting system is driven with $r(t) = u_{-2}(t)$ and the maximum error is not to exceed 0.3 radian. Determine the maximum horsepower necessary to drive this load.

6-23. The control system of Ex. 6-14-1 is redesigned in Ex. 6-16-1. This is done by using error rate damping as shown in Fig. 6-16-1. Find the value of K_f to do this.

6-24. A control system is designed as discussed in Prob. 6-23. However the signal $V_6(s)$ is inadvertently connected so that it passes through the summing point with a minus sign. Calculate the $e(t)$ response of the resulting system with $r(t) = u_{-2}(t)$.

6-25. The control system of Fig. 6-16-2 has the following constants

$$J = 15 \times 10^{-6} \text{ slug-ft}^2$$

$$D = 1.1 \times 10^{-4} \text{ ft-lb-sec}$$

$$K_1 = 0.99 \times 10^{-4} \text{ ft-lb/(radians/sec)}$$

$$K = 4.0 \times 10^{-3} \text{ ft-lb/radian}$$

The system is at rest and the input is $r(t) = u_{-2}(t)$. Find $e(t)$.

6-26. The control system of Fig. 6-18-2 has the following constants:

$$J = 15 \times 10^{-6} \text{ slug-ft}^2$$

$$D = 1.1 \times 10^{-4} \text{ ft-lb/rad/sec}$$

$$K = 4.0 \times 10^{-3} \text{ ft-lb/rad}$$

$$K_2 = 1 \times 10^{-2} \text{ ft-lb sec/rad}$$

The system is at rest and the input is $r(t) = u_{-2}(t)$. Find $e(t)$. Let K_2 be a variable. How large can K_2 become and have a stable system?

6-27. Error rate and integral compensation are combined in a single system. Find $G(s)$, $C(s)/R(s)$ and $E(s)/R(s)$ for the resulting system.

6-28. The $C(s)$ for a system with a step-function input is

$$C(s) = \frac{8,000}{s[s^4 + 24s^3 + 1,088s^2 + 4,160s + 8,000]}$$

Find $c(t)$.

6-29. Approximate the transfer function for the system of Prob. 6-28 by its dominant pole pair and find the new $c(t)$. Compare the result with the $c(t)$ found in Prob. 6-28.

6-30. The $C(s)$ for a system with a step-function input is

$$C(s) = \frac{500}{s[s^4 + 21s^3 + 521s^2 + 520s + 500]}$$

Find $c(t)$.

6-31. Approximate the transfer function for the system of Prob. 6-30 by its dominant polepair and find the new $c(t)$. Compare the result with the $c(t)$ found in Prob. 6-30.

6-32. Extend Table 6-21-2 and Table 6-21-3 to the next higher type number.

6-33. The system of Prob. 6-15 has the following function applied to the input until steady state is reached

$$r(t) = 1.5 \sin(10t + 50^\circ) - 0.8 \cos(30t - 20^\circ)$$

Find $e(t)_{ss}$ and $c(t)_{ss}$.

6-34. The system of Prob. 6-25 has the following function applied to the input steady state is reached

$$r(t) = 2.5 \cos(10t - 45^\circ) - 1.5 \sin(30t + 58^\circ)$$

Find $e(t)_{ss}$ and $c(t)_{ss}$.

6-35. Repeat Prob. 6-34 for the system of Prob. 6-26.

6-36. The system of Prob. 6-28 has the following function applied to the input until steady state is reached.

$$r(t) = 2.5 \sin(t - 30^\circ) - 1.5 \cos(2t + 50^\circ)$$

Find $e(t)_{ss}$ and $c(t)_{ss}$.

6-37. Repeat Prob. 6-36 for the system of Prob. 6-30.

6-38. For the system of Ex. 6-19-1, draw the pole pattern to scale and use the geometrical interpretation of evaluating the residues in the poles at $-3 + j3$ and at $-30 + j30$.

6-39. For the system of Prob. 6-28, draw the pole pattern to scale and use the geometrical interpretation of evaluating the residues on one of the complex poles near the origin and on one far away.

6-40. Given a system described by

$$\frac{C(s)}{R(s)} = \frac{2(s + 2.2)}{2.2(s^2 + 0.2s + 1)(s + 2)}$$

Draw the pole-zero pattern to scale and use the geometrical interpretation of determining the frequency response curve for the values of $\omega = 0.2, 0.6, 0.8, 1.0, 1.2, 1.4$, and 2.0 .

6-41. The system of Prob. 6-15 is driven at a constant angular velocity of 1.5 radians/sec until steady state is reached and the input is stopped at $t = 0$. Find $c(t)$ after $t = 0$.

6-42. The system of Prob. 6-25 is driven at a constant angular velocity of 2.5 radians/sec until steady state is reached and at $t = 0$ the input is changed to a constant angular velocity of 1.0 . Find $c(t)$ after $t = 0$. Arbitrarily choose $r(0) = 0$.

6-43. Repeat Prob. 6-42 for the system of Prob. 6-26.

6-44. The system of Prob. 6-25 is at rest when at $t = 0$ a constant torque disturbance T_1 is applied to the output in such a direction to tend to make $c(t)$ positive. Find the equation for $e(t)$ and $c(t)$.

6-45. Repeat Prob. 6-44 for the system of Prob. 6-26. Why is $e(t)_{ss}$ equal to zero in this problem but it is not in Prob. 6-44?

6-46. The coefficient K_a of Eq. 6-7-2 is to be determined in terms of the quantities as shown in Fig. 6-7-2. The potentiometer is such that it can be turned through N revolutions. Each d-c voltage has a magnitude of E as shown.

6-47. In a system similar to that of Fig. 6-16-1, the signal $v_6(t)$ is to be obtained by taking the derivative of the input signal $r(t)$ instead of the $v_1(t)$ signal as shown. The summing point is left unchanged. Analyze the resulting system. Find the value of K_f which yields zero steady-state error with a unit-ramp function input.

6-48. In Chapter 3 initial conditions are moved into sources. Develop an analogous procedure for handling the initial conditions in closed-loop systems. Rework any of the examples or problems that involve initial conditions as assigned using the procedure developed.

STABILITY

7-1. Introduction. A control system is designed to have the output follow the input as faithfully and as rapidly as possible. To achieve this a high gain around the closed loop is necessary. In most systems, however, the gain cannot be increased indefinitely without the system becoming unstable.

It is of interest to explore in a qualitative manner the way in which a physical system can become unstable before attempting a more quantitative study. Passive networks consisting of R 's, L 's, and C 's connected in any arrangement are always stable since the transient portion of a given response always decays to zero. For the hypothetical case of a network made up of pure L 's and C 's, the transient response never decays but in the practical situation, there is always resistance present.

In a closed-loop system each component can be stable, and yet the closed-loop system can be unstable. The primary reasons for this are a gain of energy inside the loop and the fact that the signal traveling around the loop usually encounters several time delays. Energy is stored in electrical and mechanical components, etc., and since the energy cannot change instantaneously from one value to another, the signal encounters the time delays.

The situation that can be explained in simple terms is the steady-state sinusoidal response. Under these conditions the gain and time delays effect a change in magnitude and phase of the signal.

To explore this situation, the system with the loop opened as shown in Fig. 7-1-1 is examined. The input $r(t)$ is given a value of zero and a sinusoidal signal of unit amplitude is applied at point E and is left long enough for the system to reach steady state. If the system is such that the feedback signal B is also of unit magnitude but is 180° out of phase with the signal E , then due to the action of the summing point, the signals E' and E are identical. Hence, the loop can be closed and the signal applied to the system at E removed with the result that the system continues

to oscillate indefinitely. Although the amplitude of the oscillations does not build up, the input to the system is no longer in control of the output and this system can be classified as absolutely unstable.

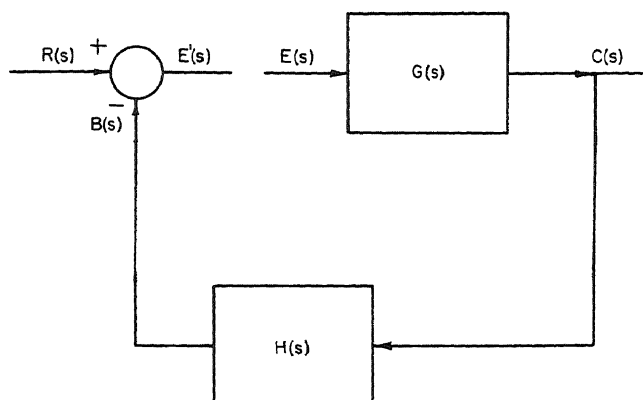


Fig. 7-1-1.

Because a control system must be stable before it can perform any of its functions, stability is a subject that should receive considerable attention. Of the numerous methods for determining stability the following methods are considered:

- (1) The characteristic polynomial is factored for its zeros.
- (2) The Routh's criterion of stability is applied to the characteristic polynomial to determine whether it has any zeros in the right half s -plane.
- (3) The frequency response of the open-loop system is plotted and the Nyquist criterion of stability is used to determine whether the closed-loop system is stable.
- (4) The Bode diagram can be interpreted to yield information about stability.
- (5) The root-locus method indicates when a closed-loop system becomes unstable as the gain of the open-loop system is increased.
- (6) The analog computer can be used to determine system stability.

Of these six methods, methods (1), (2), and (3) are explained in this chapter, method (4) in Chapter 9, method (5) in Chapter 11, and method (6) in Chapter 14.

7-2. Determining the Zeros of the Characteristic Polynomial.

The characteristic polynomial appears in the denominator of the closed-loop system. The absolute stability of the system can be determined

from the location of the zeros of the characteristic polynomial which in turn are the poles of the closed-loop transfer function. If the transfer function has poles on the $j\omega$ -axis or in the right half s -plane, the system is unstable. If all the poles are in the left half s -plane, the system is stable.

High-degree polynomials cannot be factored by formula and all known methods of factoring such polynomials involve some method of successive approximations. The Lin method as explained in Chapter 6 is typical of such methods. Not only do they demand a great deal of computational effort, but the devious relationship between the zeros and the coefficients of the polynomial obscure the engineer's efforts at design. For example, the designer factors a certain characteristic polynomial and examines the resulting poles of the closed-loop transfer function. If the set of poles obtained is not satisfactory, he has no method of determining the changes that should be made in the coefficients of the polynomial to relocate the poles in a desired manner. About the best the designer can do is to adopt a trial and error procedure.

For these reasons, other methods of design have been developed.

7-3. Routh's Test for Stability. Descartes' rule of signs gives a method by which the maximum number of positive (or negative) real zeros of a polynomial can be determined. If a zero is complex, however, this rule gives no information about the sign of the real part of complex zeros. A necessary condition for all the zeros of a polynomial to be in the left half s -plane is that all the coefficients of the polynomial be present and be positive. However, this is not a sufficient condition because under certain circumstances, all the coefficients may be present and positive and yet the polynomial may have zeros in the right half s -plane. Under these conditions the zeros in the right half plane will be complex with positive real parts. For the remainder of this discussion it is assumed (unless otherwise stated) that all the coefficients of the polynomial being considered are present and are positive.

A general method of writing a polynomial is

$$p(s) = a_n s^n + a_{n-1} s^{n-1} + \cdots + a_2 s^2 + a_1 s + a_0 \quad (7-3-1)$$

The terms of the polynomial can be rearranged so that all even terms and odd terms are grouped separately. If, for example, n in Eq. 7-3-1 is even, then $p(s)$ can be written as

$$\begin{aligned} p(s) = & (a_n s^n + a_{n-2} s^{n-2} + \cdots + a_2 s^2 + a_0) \\ & + (a_{n-1} s^{n-1} + a_{n-3} s^{n-3} + \cdots + a_3 s^3 + a_1 s) \end{aligned} \quad (7-3-2)$$

For convenience, the even part of $p(s)$ is called $m(s)$ and the odd part $n(s)$. Therefore $p(s)$ can be written as

$$p(s) = m(s) + n(s) \quad (7-3-3)$$

A polynomial that has all its zeros in the left half s -plane is called a Hurwitz polynomial. If a polynomial is Hurwitz and the ratio of $m(s)$ to $n(s)$ is formed,

$$\psi(s) = \frac{m(s)}{n(s)} \quad (7-3-4)$$

the result is a reactance function. A reactance function is a driving point impedance (or admittance) of a network containing only L 's and C 's. If Eq. 7-3-4 represents a reactance function, then $1/\psi(s) = n(s)/m(s)$ also represents a reactance function.

One method of testing a polynomial to determine whether it is Hurwitz, is to form the ratio of $m(s)$ to $n(s)$ and then test whether the resulting $\psi(s)$ can be synthesized as an L - C driving point impedance. If the $\psi(s)$ can be so synthesized, then $p(s)$ is Hurwitz. The synthesis aspect of this procedure is explained in Chapter 12.

Routh's test is a mechanistic procedure whereby a $\psi(s)$ can be tested to see whether it is a reactance function. As explained in Chapter 12, if the $\psi(s)$ tests as a reactance function, an L - C network can be determined from the coefficients of the Routh table.

The Routh table is formed by starting with the coefficients of $m(s)$ on one row the coefficients of $n(s)$ on another row. The coefficients of $m(s)$ are placed in the first row if $m(s)$ is of higher degree than $n(s)$ whereas the coefficients of $n(s)$ are placed in the first row if $n(s)$ is of higher degree than $m(s)$.

To be specific, the procedure is demonstrated on the polynomial

$$a_6s^6 + a_5s^5 + a_4s^4 + a_3s^3 + a_2s^2 + a_1s + a_0 \quad (7-3-5)$$

The coefficients of the polynomial are placed in two rows as follows:

$$\begin{array}{cccc} a_6 & a_4 & a_2 & a_0 \\ a_5 & a_3 & a_1 & \end{array} \quad (7-3-6)$$

The coefficients of the next row are formed from these two rows:

$$\begin{aligned} b_4 &= \frac{a_5a_4 - a_6a_3}{a_5} \\ b_2 &= \frac{a_5a_2 - a_6a_1}{a_5} \\ b_0 &= \frac{a_5a_0 - a_6(0)}{a_5} = a_0 \end{aligned} \quad (7-3-7)$$

The zero in the equation for b_0 represents the blank space at the end of the second row in the array of Eq. 7-3-6.

This row of coefficients is added to the array of Eq. 7-3-6, resulting in

$$\begin{array}{cccc} a_6 & a_4 & a_2 & a_0 \\ a_5 & a_3 & a_1 & \\ b_4 & b_2 & b_0 & \end{array} \quad (7-3-8)$$

The coefficients of the next row are formed from the last two rows in much the same manner as the row of b 's is formed from the first two rows.

$$\begin{aligned} c_3 &= \frac{b_4 a_3 - a_5 b_2}{b_4} \\ c_1 &= \frac{b_4 a_1 - a_5 b_0}{b_4} \end{aligned} \quad (7-3-9)$$

This row of coefficients is added to the array of Eq. 7-3-8, resulting in

$$\begin{array}{cccc} a_6 & a_4 & a_2 & a_0 \\ a_5 & a_3 & a_1 & \\ b_4 & b_2 & b_0 & \\ c_3 & c_1 & & \end{array} \quad (7-3-10)$$

The coefficients of the next row are again formed from the last two rows as

$$\begin{aligned} d_2 &= \frac{c_3 b_2 - b_4 c_1}{c_3} \\ d_0 &= \frac{c_3 b_0 - b_4(0)}{c_3} = b_0 = a_0 \end{aligned} \quad (7-3-11)$$

This process is continued and the entire Routh table is completed as

$$\begin{array}{c|cccc} s^6 & a_6 & a_4 & a_2 & a_0 \\ s^5 & a_5 & a_3 & a_1 & \\ s^4 & b_4 & b_2 & b_0 & \\ s^3 & c_3 & c_1 & & \\ s^2 & d_2 & d_0 & & \\ s^1 & e_1 & & & \\ s^0 & f_0 & & & \end{array} \quad (7-3-12)$$

The column to the left indicates the highest degree in s associated with the appropriate row.

The Routh table for a polynomial of any degree is formed in the same manner. A polynomial of degree n has a table that contains $n + 1$ rows.

If the polynomial being tested is Hurwitz, there are no changes of signs among the coefficients of the first column. If there are changes in signs in the first column, the polynomial is not Hurwitz. In addition to this, the Routh table yields the information that the number of changes in

signs in the first column is equal to the number of zeros of the polynomial in the right half s -plane.

Additional comments are best made through the use of examples.

Example 7-3-1. The polynomial to be tested is

$$s^5 + 4s^4 + 9s^3 + 44s^2 + 98s + 60 \quad (7-3-13)$$

The following Routh table is formed:

s^5	1	9	98	(7-3-14)
s^4	4	44	60	
s^3	-2	83		
s^2	210	60		
s^1	83.57			
s^0	60			

Since there are two changes of signs in the first column, the polynomial of Eq. 7-3-13 has two zeros in the right half s -plane.

It is interesting to note that if a polynomial is formed by adding the s^4 row to the s^3 row as

$$4s^4 - 2s^3 + 44s^2 + 83s + 60 \quad (7-3-15)$$

that the resulting polynomial also has two zeros in the right half plane. This can be seen because if the first row of the table of Eq. 7-3-14 is removed the remainder of the table tests this polynomial.

In a similar manner, the polynomial formed by adding the s^3 to the s^2 row as

$$-2s^3 + 210s^2 + 83s + 60 \quad (7-3-16)$$

has only one zero in the right half plane. The polynomial formed by adding rows s^2 and s^1

$$210s^2 + 83.57s + 60 \quad (7-3-17)$$

has no zeros in the right half plane.

Example 7-3-2. The polynomial to be tested is

$$s^5 + 42s^4 + 601s^3 + 3552s^2 + 7220s + 11,600 \quad (7-3-18)$$

A transformation of the variable

$$s = 10z \quad (7-3-19)$$

yields coefficients which are easier to manipulate

$$z^5 + 4.2z^4 + 6.01z^3 + 3.552z^2 + 0.722z + 0.116 \quad (7-3-20)$$

The following table is formed:

z^5	1	6.01	0.722	(7-3-21)
z^4	4.2	3.552	0.116	
z^3	5.16	0.694		
z^2	2.98	0.116		
z^1	0.49			
z^0	0.116			

Since there are no changes of signs in the first column, the polynomial of Eq. 7-3-20 (and also 7-3-18) has no zeros in the right half plane and is therefore Hurwitz.

It is interesting to note that any two adjacent rows throughout this array may be added together and the result is a Hurwitz polynomial. For example, if the rows identified by z^3 and z^2 are added, yielding

$$5.16z^3 + 2.98z^2 + 0.694z + 0.116 \quad (7-3-22)$$

the resulting polynomial can be seen to be Hurwitz because the table of Eq. 7-3-21 from row z^3 down tests this polynomial. Since there are no changes of signs in the first column, the polynomial of Eq. 7-3-22 is Hurwitz.

From this line of reasoning, it can be seen that if a polynomial is Hurwitz, all the coefficients in the corresponding Routh table must be positive. If a polynomial is not Hurwitz, a negative sign that appears anywhere in the table will eventually work its way over to the first column as successive rows are formed. Therefore, if the designer of a control system is testing a polynomial only to determine whether it is Hurwitz, he may stop as soon as a negative sign appears anywhere in the Routh table, because the polynomial is not Hurwitz. However, if he wants to know the number of zeros in the right half s -plane, the table should be completed so that the first column is available for inspection.

Since the sum of any two adjacent rows produces a Hurwitz polynomial, the ratio of any two adjacent rows forms a reactance function. As explained in Chapter 12, a reactance function has simple poles and zeros on the $j\omega$ -axis in the s -plane and the poles and zeros are interlaced. This means that the zeros of any row of a Routh table obtained from a Hurwitz polynomial are interlaced with the zeros from the adjacent rows. Since any row can be multiplied by any constant without changing the location of its zeros, it can be seen that any row in the Routh table may be multiplied by any constant. Therefore, the division by a_5 (for example) as given by the relationships expressed in the array of Eq. 7-3-7, is not a necessity as any other number would do equally well. In general, the division by this number serves a useful purpose. For one thing this division keeps the coefficient of the Routh table from becoming unwieldy. Even more important, when the driving point impedance of a reactance functions is to be synthesized, the Routh table can be used directly if the table is formed in this manner. This procedure is explained in Chapter 12.

Example 7-3-3. The polynomial to be tested is

$$s^5 + 2s^4 + 3s^3 + 6s^2 + 4s + 5 \quad (7-3-23)$$

The following table is formed:

s^5	1	3	4	(7-3-24)
s^4	2	6	5	
s^3	0	1.5		

As soon as the third row is formed, a zero appears. If the original polynomial were Hurwitz, then any two adjacent rows throughout the remainder of the table when added should form a Hurwitz polynomial. When a polynomial is formed by adding the second and third rows, a term is missing; therefore, this polynomial is not Hurwitz. Finally it can be reasoned that the original polynomial cannot be Hurwitz either.

Whenever a zero appears in a row and the remaining coefficients of the row are not zero the original polynomial is not Hurwitz. If the designer wishes to determine the number of zeros of the polynomial that are in the right half s -plane, he can proceed in one of several ways. One of the most convenient is to form a polynomial by making the substitution

$$s = \frac{1}{z} \quad (7-3-25)$$

This substitution effectively reverses the arrangement of the coefficients and in this case yields the polynomial

$$5z^5 + 4z^4 + 6z^3 + 3z^2 + 2z + 1 \quad (7-3-26)$$

Because of the substitution, the zeros of the polynomial of Eq. 7-3-26 are the reciprocals of the zeros of the polynomial of Eq. 7-3-23. Since the reciprocal of a zero with a positive real part also has a positive real part, the new polynomial has the same number of zeros in the right half plane as does the original.

To test the new polynomial the table is formed:

z^5	5	6	2
z^4	4	3	1
z^3	2.25	0.75	
z^2	1.66	1	
z^1	-0.6		
z^0	1		

Therefore, the new and original polynomials have two zeros in the right half plane.

In certain situations when the substitution $s = 1/z$ is made, the resulting table also has a zero in the first column. Therefore other procedures are required. One procedure is to replace the zero in the table of Eq. 7-3-24 by an arbitrarily small number and to finish the table.

s^5	1	3	4
s^4	2	6	5
s^3	$\begin{cases} 0 \\ \epsilon \end{cases}$	1.5	
s^2	$6 - \frac{3}{\epsilon}$	5	
s^1	$1.5 - \frac{5\epsilon^2}{6\epsilon - 3}$		
s^0	5		

An examination of the first column shows there are two changes in signs regardless of whether ϵ is assumed positive or negative.

Example 7-3-4. The polynomial to be tested is

$$s^6 + 3s^5 + 4s^4 + 6s^3 + 9s^2 + 15s + 10 \quad (7-3-27)$$

The following table is formed

s^6	1	4	9	10	(7-3-28)
s^5	3	6	15		
s^4	2	4	10		
s^3	0	0			

The entire fourth row is composed of zeros. An examination of the manner in which the coefficients are formed indicates that the coefficients of the previous two rows are directly proportional. This in turn results from the fact that these two rows contain a common factor. To demonstrate this, the second and third row respectively are written as

$$3s^5 + 6s^3 + 15s = 3s(s^4 + 2s^2 + 5) \quad (7-3-29)$$

$$2s^4 + 4s^2 + 10 = 2(s^4 + 2s^2 + 5) \quad (7-3-30)$$

The common factor between the two rows is

$$s^4 + 2s^2 + 5 \quad (7-3-31)$$

This common factor is also found in the first and second rows and in the original polynomial as can be seen from the following development. The even function $m(s)$ and the odd function $n(s)$ which make up the original polynomial can be written as

$$\begin{aligned} m(s) &= m_1(s)c(s) \\ n(s) &= n_1(s)c(s) \end{aligned} \quad (7-3-32)$$

where $c(s)$ is the common factor. The polynomial $p(s)$ can be written as

$$p(s) = m(s) + n(s) = m_1(s)c(s) + n_1(s)c(s) = [m_1(s) + n_1(s)]c(s) \quad (7-3-33)$$

In this specific example, the polynomial of Eq. 7-3-27 has the polynomial of Eq. 7-3-31 as a factor and can be written as

$$p(s) = (s^2 + 3s + 2)(s^4 + 2s^2 + 5) \quad (7-3-34)$$

The Routh table down to the row of zeros tests the $m_1(s) + n_1(s)$ factor which in this specific example is $s^2 + 3s + 2$. Therefore, $s^2 + 3s + 2$ is Hurwitz since there are no changes of signs in the first column down to the zeros. However, the $c(s) = s^4 + 2s^2 + 5$ factor remains to be tested.

The common factor $c(s)$ is always an even function which in general can be expressed as

$$c(s) = c_n s^n + c_{n-2} s^{n-2} + \cdots + c_2 s^2 + c_0 \quad (7-3-35)$$

where n is even. To continue the discussion, the substitution

$$z = s^2 \quad (7-3-36)$$

is made, yielding

$$c(z) = c_n z^{n/2} + c_{n-2} z^{(n-2)/2} + \cdots + c_2 z + c_0 \quad (7-3-37)$$

If $c(z)$ has all of its zeros on the negative real axis, then $c(s)$ has all of its zeros on the $j\omega$ -axis. However, if $c(z)$ has a complex pair of zeros

$$z_1 \quad \text{and} \quad z_2 = re^{+ja}, re^{-ja} \quad (7-3-38)$$

then $c(s)$ has four zeros

$$s_1, s_2, s_3, \quad \text{and} \quad s_4 = \pm\sqrt{re^{+ja/2}}, \pm\sqrt{re^{-ja/2}} \quad (7-3-39)$$

two of which are in the right half s -plane and the other two in the left half s -plane.

In general, the zeros of $c(s)$ occur in pairs located radially through the origin from each other with the result that the zero pattern is symmetrical with respect to the origin. The only possible way in which an even function such as $c(s)$ can be classed as a Hurwitz polynomial is to extend the definition of the left half s -plane so that it also includes the $j\omega$ -axis and to have a $c(s)$ polynomial with all its zeros on the $j\omega$ -axis. Therefore, when a row of zeros appears in the Routh table, the best result to be expected is for the zeros to be on the $j\omega$ -axis. For control systems, the corresponding transfer function has poles on the $j\omega$ -axis and the system can be classified as unstable. However in certain situation it is desirable to complete the table and determine whether the zeros are on the $j\omega$ -axis. For example, this technique is used in the next article.

What is needed to complete the table of Eq. 7-3-28 is a procedure for testing an even polynomial to determine whether its zeros are on the $j\omega$ -axis.

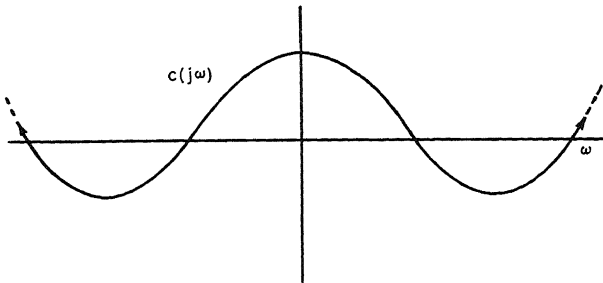


Fig. 7-3-1.

If an even polynomial has its zeros on the $j\omega$ -axis, $c(j\omega)$ vs. ω will plot as indicated in Fig. 7-3-1. The derivative of this function has zeros that are interlaced with those of $c(j\omega)$. Therefore

$$c(s) + \frac{d}{ds} c(s) \quad (7-3-40)$$

forms a Hurwitz polynomial and can be tested by the Routh table.

In the particular example being considered, the common factor is

$$c(s) = s^4 + 2s^2 + 5 \quad (7-3-41)$$

and its derivative is

$$\frac{dc(s)}{ds} = 4s^3 + 4s \quad (7-3-42)$$

The coefficients from these polynomials are entered as the first two rows of a Routh table and the table is completed. The coefficients of the polynomial of Eq. 7-3-42 are divided by 4 for convenience before entering the coefficients in the second row.

s^4	1	2	5	
s^3	1	1		
s^2	1	5		
s^1	-4			
s^0	5			

(7-3-43)

Since there are two changes of signs in the first column, the polynomial of Eq. 7-3-41 and also the original polynomial of Eq. 7-3-27 have two zeros in the right half plane.

Example 7-3-5. The polynomial to be tested is

$$s^6 + 3s^5 + 6s^4 + 12s^3 + 11s^2 + 9s + 6 \quad (7-3-44)$$

The following table is formed

s^6	1	6	11	6	
s^5	3	12	9		
s^4	2	8	6		
s^3	$\begin{Bmatrix} 0 & 0 \\ 1 & 2 \end{Bmatrix}$				
s^2	4	6			
s^1	0.5				
s^0	6				

(7-3-45)

The table proceeds in the usual manner until a row of zeros is encountered in the fourth row. The common factor is determined from the third row as

$$c(s) = s^4 + 4s^2 + 3 \quad (7-3-46)$$

The derivative of this is

$$\frac{dc(s)}{ds} = 4s^3 + 8s \quad (7-3-47)$$

These coefficients are divided by 4 and are substituted for the row of zeros. The table again proceeds in the usual manner.

The fact that there is no change of signs in the first column down to the row of zeros indicates that the polynomial that multiplies $s^4 + 4s^2 + 3$, to complete the original polynomial, is Hurwitz. The fact that there is no change of signs in the remainder of the first column indicates that the even polynomial has all its zeros on the $j\omega$ -axis.

Example 7-3-6. This example is added to demonstrate a method of testing a polynomial to determine whether it has all negative real zeros. The polynomial used for this purpose is

$$s^3 + 7s^2 + 14s + 8 \quad (7-3-48)$$

The substitution

$$s = z^2 \quad (7-3-49)$$

is made and the polynomial becomes

$$z^6 + 7z^4 + 14z^2 + 8 \quad (7-3-50)$$

If the polynomial in z has all its zeros on the $j\omega$ -axis, the polynomial in s has all its zeros on the negative real axis. To test the polynomial in z , it is differentiated and divided by 2 as

$$3z^5 + 14z^3 + 14z \quad (7-3-51)$$

The polynomials of Eqs. 7-3-50 and 7-3-51 form the first two rows of a Routh table

z^6	1	7	14	8
z^5	3	14	14	
z^4	2.33	9.33	8	
z^3	2	3.71		
z^2	5	8		
z^1	0.51			
z^0	8			

(7-3-52)

Since there are no changes of signs in the first column, the polynomial in z has all its zeros on the $j\omega$ -axis and the original polynomial in s has all its zeros on the negative real axis.

Example 7-3-7. In Art. 6-18, integral compensation is discussed and the characteristic polynomial for the system is

$$Js^3 + Ds^2 + Ks + K_2 \quad (7-3-53)$$

The corresponding Routh table is

s^3	J	K
s^2	D	K_2
s^1	$\frac{DK - JK_2}{D}$	
s^0	K_2	

(7-3-54)

The first column of this table is positive as long as

$$DK > JK_2 \quad (7-3-55)$$

This relationship is used in Chapter 6 as Eq. 6-18-10.

7-4. Routh's Test Used to Determine Relative Stability. In a control system, Routh's test is usually applied to the characteristic polynomial which appears in the denominator of the closed-loop transfer function. The result is that the closed-loop system is tested for the existence of poles in the right half s -plane. Routh's criterion in its simple form tests only for absolute stability and yields no information about relative stability.

If the Routh criterion is combined with a technique of shifting the $j\omega$ -axis, much more information can be determined. This procedure is similar to the manner in which Horner's method is used to find a real zero. Routh's test is first applied to the characteristic polynomial as given.

If no change of signs occurs in the first column, the $j\omega$ -axis is shifted to the left and the Routh test is repeated. If there are still no changes of sign in the first column, the $j\omega$ -axis is again shifted to the left. After a few tries, the zeros of the polynomial farthest to the right can be located in a vertical strip in the s -plane. By dividing this strip into substrips the real part of the complex zeros can be found to any desired degree of accuracy.

To illustrate the method, the polynomial used in Ex. 7-3-2 is again considered. This polynomial is

$$s^5 + 42s^4 + 601s^3 + 3552s^2 + 7220s + 11,600 \quad (7-4-1)$$

Again the change of variables

$$s = 10z \quad (7-4-2)$$

is made yielding the polynomial in z

$$z^5 + 4.2z^4 + 6.01z^3 + 3.552z^2 + 0.722z + 0.116 \quad (7-4-3)$$

The polynomial as given is tested in the table of Eq. 7-3-21 and is found to be Hurwitz. Therefore the j -axis is to be shifted to the left and the test repeated. After a few tries, the j -axis is finally shifted to -0.1 . The new polynomial is found by the technique used in Horner's method as

1	4.2	6.01	3.552	0.722	0.116	-0.1
	-0.1	-0.41	-0.560	-0.2992	-0.04228	
1	4.1	5.60	2.992	0.4228	0.07372*	
	-0.1	-0.4	-0.52	-0.2472		
1	4.0	5.2	2.472	0.1756*		
	-0.1	-0.39	-0.418			
1	3.9	4.81	1.991*			
	-0.1	-0.38				
1	3.8	4.43*				
	-0.1					
1*	3.7*					

(7-4-4)

The new polynomial is obtained from the coefficients marked by the asterisk as:

$$z^5 + 3.7z^4 + 4.43z^3 + 1.991z^2 + 0.1756z + 0.07372 \quad (7-4-5)$$

This polynomial is tested by the following Routh table.

z^5	1	4.43	0.1756
z^4	3.7	1.991	0.07372
z^3	3.892	0.1557	
z^2	1.843	0.07372	
z^1	0		

(7-4-6)

All the coefficients of the fifth row of this table are zeros, therefore

$$1.843z^2 + 0.07372 \quad (7-4-7)$$

is a factor of the polynomial of Eq. 7-4-5. From this, the zeros

$$z_1 \text{ and } z_2 = \pm j0.2 \quad (7-4-8)$$

are determined. Therefore, two of the zeros of the original polynomial in z (before the axis is shifted) are

$$z_1 \text{ and } z_2 = -0.1 \pm j0.2 \quad (7-4-9)$$

and two of the zeros of the original polynomial in s are

$$s_1 \text{ and } s_2 = -1 \pm j2 \quad (7-4-10)$$

As a by-product of this procedure, the designer has available another method of factoring a polynomial. From the zeros in Eq. 7-4-10, it can be seen that the polynomial

$$s^2 + 2s + 5 \quad (7-4-11)$$

is a factor of the original polynomial in s . After a division process, this polynomial can be written as

$$(s^2 + 2s + 5)(s^3 + 40s^2 + 516s + 2320) \quad (7-4-12)$$

and the polynomial

$$s^3 + 40s^2 + 516s + 2320 \quad (7-4-13)$$

can now be factored in a similar manner. After a few tries, the $j\omega$ -axis is shifted to -10 .

1	40	516	2320	-10
	-10	-300	-2160	
1	30	216	160	
	-10	-200		
1	20	16		
	-10			
1	10			

(7-4-14)

The new polynomial is

$$s^3 + 10s^2 + 16s + 160 \quad (7-4-15)$$

This polynomial is tested by the following Routh table

s^3	1	16
s^2	10	160
s^1	0	

(7-4-16)

All the coefficients of the third row are zero, therefore

$$10s^2 + 160 \quad (7-4-17)$$

must be a factor of the polynomial of Eq. 7-4-15. From this the zeros

$$s_1 \text{ and } s_2 = \pm j4 \quad (7-4-18)$$

are determined. The zeros of the polynomial of Eq. 7-4-13 are

$$s_1 \text{ and } s_2 = -10 \pm j4 \quad (7-4-19)$$

From these zeros it can be seen that the polynomial

$$s^2 + 20s + 116 \quad (7-4-20)$$

is a factor of the polynomial of Eq. 7-4-13 which after a division process can be written as

$$(s^2 + 20s + 116)(s + 20) \quad (7-4-21)$$

The original polynomial in s has now been factored into the following zeros

$$\begin{aligned} s_1 \text{ and } s_2 &= -1 \pm j2 \\ s_3 \text{ and } s_4 &= -10 \pm j4 \\ s_5 &= -20 \end{aligned} \quad (7-4-22)$$

Although this technique adds to the usefulness of Routh's criterion, it does not overcome the devious relationship between the coefficients of a polynomial and the zeros of this polynomial. Therefore Routh's test is a useful tool but the designer must supplement it by other procedures.

7-5. Conformal Mapping. The Nyquist criterion of stability essentially invokes mapping. Before beginning an examination of Nyquist's methods some of the more elementary aspects of mapping are considered.

The functional relationship

$$y = f(x) = x^2 \quad (7-5-1)$$

can be displayed graphically on one curve if x is allowed to take on only real values. Such a graph is shown in Fig. 7-5-1.

When the variable is complex such as

$$s = \sigma + j\omega \quad (7-5-2)$$

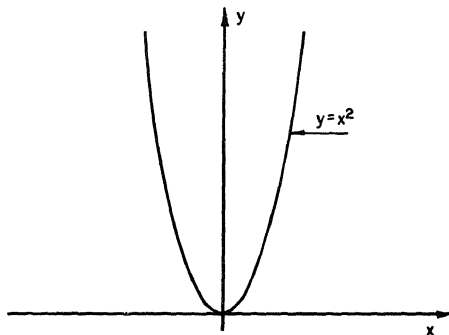


Fig. 7-5-1.

the function of the complex variable is also complex as

$$F(s) = U(\sigma, \omega) + jV(\sigma, \omega) \quad (7-5-3)$$

Where $U(\sigma, \omega)$ and $V(\sigma, \omega)$ are real functions of σ and ω . It is not possible to display this functional relationship by a single curve as is done for the real variable in Fig. 7-5-1. However, this relationship can be displayed by using two separate planes. The s -plane displays the values over which s is to range. Corresponding to each point (σ, ω) in the s -plane there is

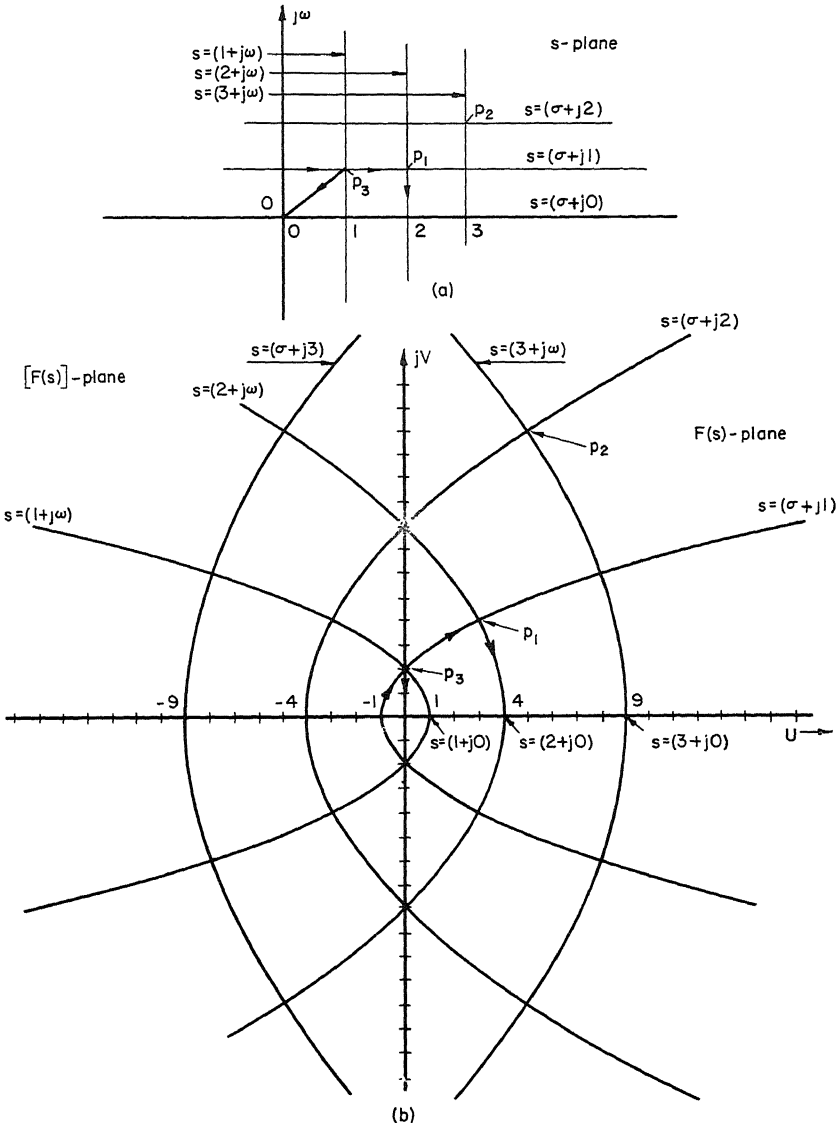


Fig. 7-5-2.

a point (U, V) in the $F(s)$ plane. The function $F(s)$ is said to map points from the s -plane to the $F(s)$ plane. Usually more information is conveyed if curves or regions in the s -plane are mapped into the $F(s)$ plane. If the mapping function is analytic, then the term conformal mapping is used. The properties of conformal mapping are explained as the discussion proceeds.

The functional relationship

$$F(s) = s^2 \quad (7-5-4)$$

is used as an example. The s -plane is shown in Fig. 7-5-2a and the $F(s)$ plane is shown in Fig. 7-5-2b. The point $s = 1 + j0$ in the s -plane maps upon the $F(s)$ plane as $F(s) = (1 + j0)^2 = 1 + j0$. The point $s = 2 + j0$ maps as $F(s) = (2 + j0)^2 = 4 + j0$. In general, the real axis σ in the s -plane maps upon the $F(s)$ plane as $F(s) = (\sigma + j0)^2 = \sigma^2 + j0$ which is a curve on the positive real axis in the $F(s)$ plane.

The line $s = 1 + j\omega$ in the s -plane maps upon the $F(s)$ plane as $F(s) = (1 + j\omega)^2 = 1 - \omega^2 + j2\omega$. The line $s = 3 + j\omega$ maps as $F(s) = (3 + j\omega)^2 = 9 - \omega^2 + j6\omega$. The line $s = \sigma + j1$ maps as $F(s) = (\sigma + j1)^2 = \sigma^2 - 1 + j2\sigma$. The line $s = \sigma + j2$ maps as $F(s) = (\sigma + j2)^2 = \sigma^2 - 4 + j4\sigma$. Other lines can be mapped in a similar manner.

In the following discussion, it is assumed a line l_1 in the s -plane maps as a line L_1 in the $F(s)$ plane and a line l_2 maps as L_2 . The two lines l_1 and l_2 intersect at a point s_o in the s -plane. If the mapping is such that the lines L_1 and L_2 intersect in the $F(s)$ plane so that the angle in both magnitude and sense between L_1 and L_2 is the same as the angle between l_1 and l_2 , the mapping is said to be conformal at the point. For the mapping to be conformal this preservation of angle must be true for every pair of lines that intersect at the point s_o .

Mapping is conformal at each point where $F(s)$ is analytic and where $F'(s)$ does not equal zero. A point at which $F'(s) = 0$ is called a critical point. The mapping function $F(s) = s^2$ is analytic at every point in the s -plane; however, the derivative is equal to zero at $s = 0$. Therefore mapping by this function is conformal at every point in the s -plane except at $s = 0$.

These concepts are explored by examining Fig. 7-5-2. In the s -plane, a point p_1 is located at the intersection of the two lines $s = 2 + j\omega$ and $s = \sigma + j1$. The point p_1 is approached from left to right along the $s = \sigma + j1$ line and is left by turning to the right through 90° along the $s = 2 + j\omega$ line. This path is indicated by arrows in Fig. 7-5-2a. In the $F(s)$ -plane shown in Fig. 7-5-2b the mapping of p_1 is labeled P_1 and the mapping of $s = \sigma + j1$ and $s = 2 + j\omega$ lines is shown. The point P_1 is approached along a line and is left by turning to the right through 90° as indicated by the arrows. Since the mapping is conformal, the magnitude and the sense of the angle are maintained. A similar type of discussion can be repeated for the point p_2 .

In the s -plane, the point p_3 is approached and left by paths that have an angle of 45° between them. The mapping of p_3 is P_3 and the paths are shown in the $F(s)$ -plane. Here again the magnitude and sense of the angle are maintained.

Since $s = 0$ is a critical point, the mapping at this point is not conformal. A radial line from the origin in the s -plane can be expressed as

$$s = \rho e^{j\alpha} \quad (7-5-5)$$

where α is a constant and ρ a variable. This line is mapped into a $F(s)$ -plane as

$$F(s) = s^2 = \rho^2 e^{j2\alpha} \quad (7-5-6)$$

Therefore at $s = 0$, the magnitude of the angle is not maintained and the mapping is not conformal.

7-6. Nyquist Criterion of Stability. Transient methods do not lend themselves to system design because of the devious relationship between the coefficients of the characteristic equation and zeros of this equation. For these reasons, designers have turned their attention to earlier work done by Nyquist and Bode, among others, in the field of feedback amplifiers. With modifications, particularly in terminology, these methods have and still do play an important part in the design of control systems.

Nyquist criterion is essentially a graphical method involving mapping. From the Nyquist diagrams the following information can be obtained:

1. Whether or not the system is absolutely stable.
2. If the system is stable, the diagrams also yield information about relative stability.
3. If the system response is not satisfactory, the diagram gives information as to how the system can be changed to improve the response.
4. If the open-loop system is stable, the Nyquist diagram can be obtained experimentally. From this experimental curve the information just discussed in Items 1, 2, and 3 can be obtained.

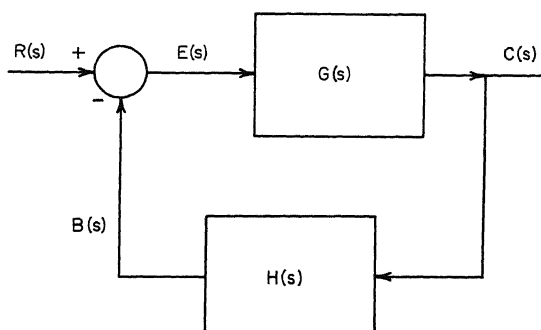


Fig. 7-6-1.

Items 1 and 2 are discussed in this chapter and Items 3 and 4 are discussed in Chapter 8.

The block diagram for a single-loop system is shown in Fig. 7-6-1. The closed-loop transfer function of this system is given by

$$\frac{C(s)}{R(s)} = \frac{G(s)}{1 + H(s)G(s)} \quad (7-6-1)$$

For the remainder of this discussion, attention is centered on the denominator of this closed-loop transfer function,

$$1 + H(s)G(s) \quad (7-6-2)$$

Zeros of this denominator are poles of the closed-loop transfer function. If the denominator has a zero in the right half s -plane, the closed-loop system has a pole in the right half s -plane and the system is unstable. Therefore any test that can detect the number of zeros of $1 + H(s)G(s)$ in the right half s -plane can be used as a test for absolute stability.

$G(s)$ is formed by multiplying the individual transfer functions in the forward loop and can be written as

$$G(s) = \frac{N_G(s)}{D_G(s)} \quad (7-6-3)$$

In a similar manner $H(s)$ is formed by multiplying the individual transfer functions in the feedback loop and can be written as

$$H(s) = \frac{N_H(s)}{D_H(s)} \quad (7-6-4)$$

With this notation, the denominator of the closed-loop transfer function becomes

$$\begin{aligned} 1 + H(s)G(s) &= 1 + \frac{N_H(s)}{D_H(s)} \frac{N_G(s)}{D_G(s)} \\ &= \frac{D_H(s)D_G(s) + N_H(s)N_G(s)}{D_H(s)D_G(s)} \end{aligned} \quad (7-6-5)$$

The numerator of this equation

$$D_H(s)D_G(s) + N_H(s)N_G(s) \quad (7-6-6)$$

is the characteristic polynomial of the closed-loop system. If this numerator has any zeros in the right half s -plane, the closed-loop system is unstable.

The numerator and denominator of Eq. 7-6-5 can be factored and put in the form.

$$1 + H(s)G(s) = \frac{K(s + z_1)(s + z_2)(s + z_3) \cdots (s + z_m)}{(s + p_1)(s + p_2)(s + p_3) \cdots (s + p_n)} \quad (7-6-7)$$

where $-z_1, -z_2, \dots, -z_m$ are the zeros of this function and $-p_1, -p_2, \dots, -p_n$ are the poles.

It is worth noting that the poles of $H(s)G(s)$ are the poles of $1 + H(s)G(s)$.

In order to explain the Nyquist diagram, it is first assumed that the location of zeros and poles of $1 + H(s)G(s)$ is known and the effect upon mapping from the s -plane to the $[1 + H(s)G(s)]$ -plane is then analyzed. Based on the information gained from this analysis, the Nyquist diagram is used in the reverse direction. That is, the location of the poles is known and from the Nyquist diagram, the location of the zeros is determined.

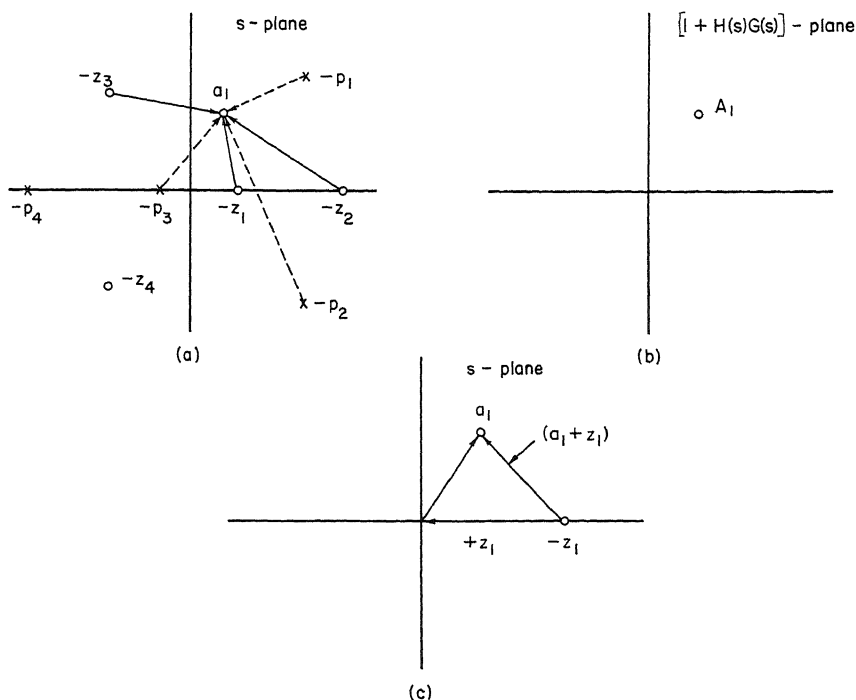


Fig. 7-7-1.

7-7. Mapping from the s -Plane to the $[1 + H(s)G(s)]$ -Plane.

The assumed locations of a few of the zeros and the poles of $1 + H(s)G(s)$ in the s -plane are shown in Fig. 7-7-1a. A point a_1 in the s -plane maps into the point A_1 in the $[1 + H(s)G(s)]$ -plane as

$$1 + H(a_1)G(a_1) = \frac{(a_1 + z_1)(a_1 + z_2) \cdots (a_1 + z_m)}{(a_1 + p_1)(a_1 + p_2) \cdots (a_1 + p_n)} = A_1 \quad (7-7-1)$$

Without a complete set of numerical values for the z 's and p 's it is not possible to determine the exact value for A_1 ; however, the point A_1 exists somewhere in the $[1 + H(s)G(s)]$ -plane and such a point is suggested in Fig. 7-7-1b.

A visualization of the manner in which A_1 is determined from the poles and zeros is helpful in the discussions that follow. A phasor from $-z_1$ to the origin is numerically equal to z_1 and is shown in Fig. 7-7-1c. A phasor from the origin to a_1 is numerically equal to a_1 . The phasor sum of $a_1 + z_1$ is the phasor from $-z_1$ to the point a_1 as indicated in the figure.

Similarly, the other factors in the numerator of Eq. 7-7-1 can be interpreted as phasors. The factor $a_1 + z_2$ is a phasor from $-z_2$ to a_1 and the factor $a_1 + z_m$ is a phasor from $-z_m$ to a_1 . A few of these phasors are shown with solid lines in Fig. 7-7-1a.

The factors in the denominator can be interpreted as phasors in an identical manner. The factor $a_1 + p_1$ is a phasor from $-p_1$ to a_1 and the factor $a_1 + p_n$ is a phasor from $-p_n$ to a_1 . A few of these phasors are shown with dashed lines in Fig. 7-7-1a.

The phasor product of the factors that appear in the numerator of Eq. 7-7-1 divided by the phasor product of the factors that appear in the denominator is equal to A_1 .

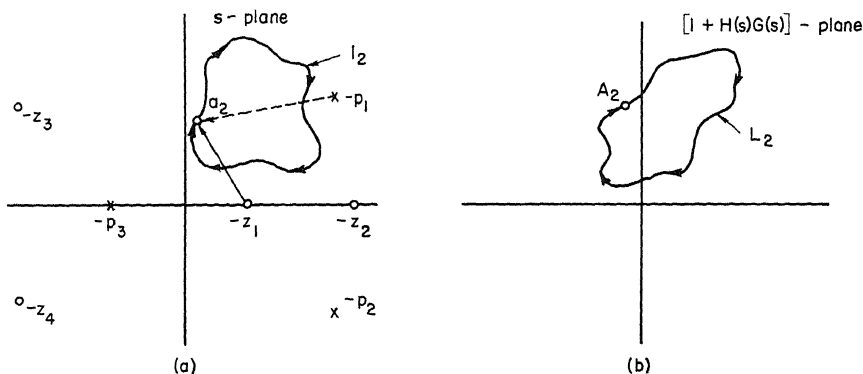


Fig. 7-7-2.

The visualization of phasors is very helpful when a contour from the s -plane is mapped into the $[1 + H(s)G(s)]$ -plane. For example, in Fig. 7-7-2a, the point a starts at a_2 and moves around a contour l_2 in the s -plane. The point a maps into A . Correspondingly, the point A starts at A_2 and traverses the contour L_2 thus mapping l_2 into L_2 . In this example, the contour l_2 does not enclose any zeros or poles. It is not possible to predict the exact shape of L_2 but certain of its characteristics can be determined.

As the point a moves around l_2 , the phasor $a + z_1$ starts as $a_2 + z_1$ and swings through a set of values returning to $a_2 + z_1$ as a returns to a_2 . Therefore, the phasor $a + z_1$ has no net change in magnitude or angle as a traverses the contours l_2 . Similarly, it can be reasoned that the phasor from all the other zeros and poles will have no net change in magnitude

or angle. Therefore, as A traverses L_2 , there is no net change in magnitude and angle and the contour L_2 must be such that the origin is not enclosed in L_3 .

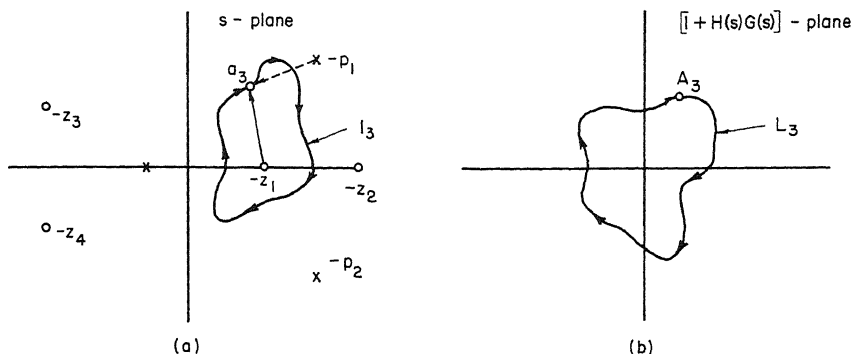


Fig. 7-7-3.

In Fig. 7-7-3a, point a starts at a_3 and traverses contour l_3 that encloses the zero at $-z_1$. Correspondingly, in Fig. 7-7-3b the point A starts at A_3 and traverses the contour L_3 . The phasor $a + z_1$, starts at $a_3 + z_1$ and undergoes a net change in phase of 360° in the clockwise direction. The phasors from all the other zeros and all the poles undergo no net change in phase. Therefore, the phasor A undergoes a net change in phase of 360° in the clockwise direction. The exact shape of the contour L_3 is not known but it must enclose the origin as shown in Fig. 7-7-3b.

Before continuing the discussion, a slightly different form of mapping is explored. Instead of mapping from the s -plane to the $[1 + H(s)G(s)]$ -plane, it is usually more convenient to map to the $[H(s)G(s)]$ -plane. The contours in these two planes only differ by 1. For example, if the contours already discussed in Figs. 7-7-2b and 7-7-3b were shifted to the left by one unit, the resulting curves are the mappings into the $(H(s)G(s))$ -plane. Therefore the point of interest is no longer whether the contour encloses the origin but rather whether the contour encloses the -1 point.

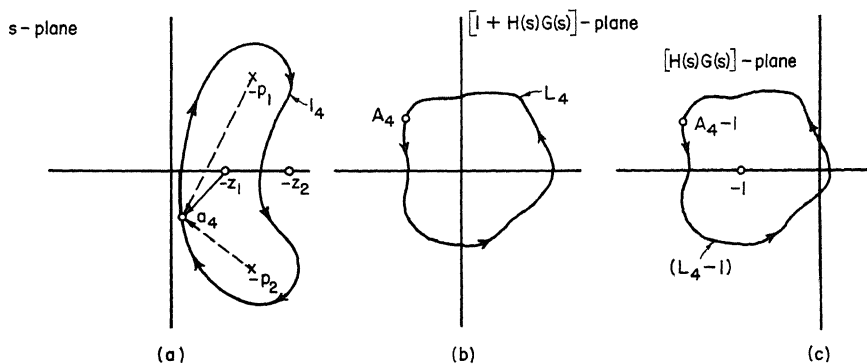


Fig. 7-7-4.

In Fig. 7-7-4a point a in the s -plane starts at a_4 and traverses contour l_4 that encloses the zero at $-z_1$ and the poles at $-p_1$ and $-p_2$. Correspondingly in the $[1 + H(s)G(s)]$ -plane of Fig. 7-7-4b, the point A starts at A_4 and traverses the contour L_4 . In the $[H(s)G(s)]$ -plane of Fig. 7-7-4, the point a_4 maps into $(A_4 - 1)$ and the contour l_4 maps into $(L_4 - 1)$. In the s -plane, each of the phasors $a + z_1$, $a + p_1$, and $a + p_2$ undergo a net change in phase of 360° in the clockwise direction whereas the phasors from all the other zeros and poles undergo no net change in phase. In the equation for A , repeated as

$$\frac{(a + z_1)(a + z_2) \cdots (a + z_m)}{(a + p_1)(a + p_2) \cdots (a + p_n)} = A \quad (7-7-2)$$

the phasor $a + z_1$ is in the numerator, whereas the phasors $a + p_1$ and $a + p_2$ are in the denominator. Therefore, the phasor A undergoes a net change in phase of 360° in the counterclockwise direction. The exact shape of the contour L_4 is not known, but it must enclose the origin as shown in Fig. 7-7-4b, and the contour $L_4 - 1$ must enclose the -1 point as shown in Fig. 7-7-4c.

The foregoing discussion can be generalized to obtain the equation

$$N = P - Z \quad (7-7-3)$$

The symbols in this equation have the following meaning:

N is the number of counterclockwise rotations of the contour in the $[H(s)G(s)]$ -plane about the point $-1 + j0$.

P is the number of poles of $H(s)G(s)$ enclosed by the contour in the s -plane. The poles of $1 + H(s)G(s)$ are the poles of $H(s)G(s)$.

Z is the number of zeros of the characteristic equation of the closed-loop system that are enclosed by the contour in the s -plane.

Eq. 7-7-3 is developed under the assumption that the contour in the s -plane is always traversed in the clockwise direction. Probs. 7-9 and 7-10 demonstrate that the choice of directions used to develop Eq. 7-7-3 are arbitrary and that other conventions can be used. However, when a set of directions is chosen, this set must be used consistently.

Eq. 7-7-3 has been developed assuming knowledge about the location of the poles and zeros of $1 + H(s)G(s)$. However, the equation is used in the reverse direction. That is, the mapping in the $[H(s)G(s)]$ -plane is used to determine N . The constant P is determined from the denominator of $H(s)G(s)$. The constant Z is determined by using Eq. 7-7-3. The details of this procedure are explained by several examples.

Example 7-7-1. For the first example, the following $G(s)$ and $H(s)$ are assumed

$$G(s) = \frac{1}{s(s+1)} \quad ; \quad H(s) = 1 \quad (7-7-4)$$

The Nyquist diagram is used first to determine absolute stability. The l_1 contour in the s -plane of Fig. 7-7-5a is chosen so as to enclose the major portion of the

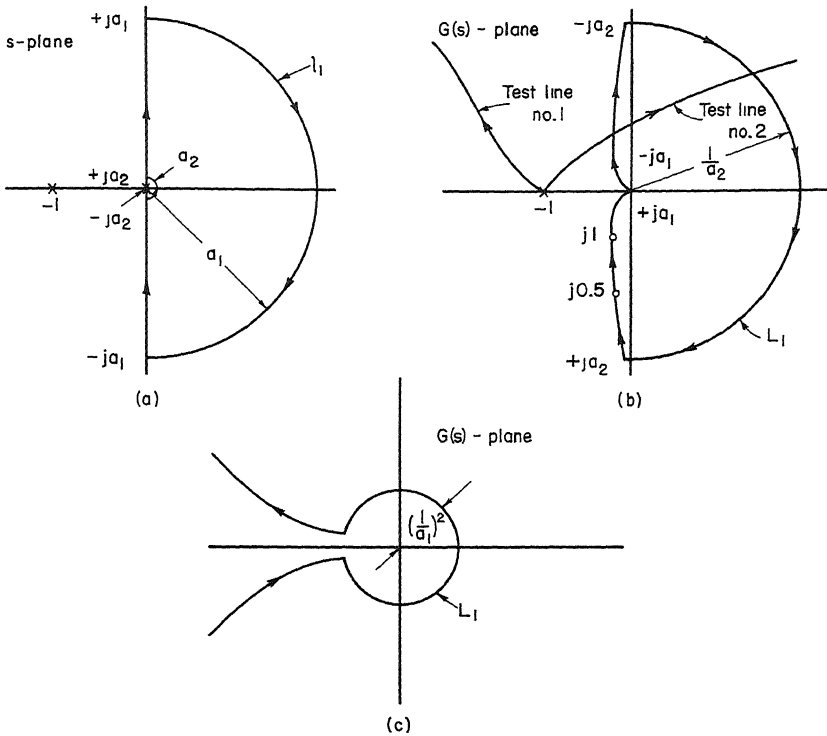


Fig. 7-7-5.

right half plane. The contour follows the $j\omega$ -axis in the large but avoids all poles on the $j\omega$ -axis. In this example, the only such pole is at the origin and l_1 follows a semi-circle of radius a_2 to avoid this pole. The l_1 contour is completed by the semi-circle of radius a_1 . By making a_2 sufficiently small and a_1 sufficiently large, the l_1 contour encloses the major portions of the right half s -plane.

The l_1 contour maps into the L_1 contour in the $[G(s)]$ - plane. The L_1 contour is developed in sections. The positive $j\omega$ -axis is mapped as follows. The s in $G(s)$ is replaced by $j\omega$,

$$G(j\omega) = \frac{1}{j\omega(j\omega + 1)} \quad (7-7-5)$$

Table 7-7-1 is formed as an aid in developing $G(j\omega)$. Since a_2 is assumed to be very small, ja_2 in the first row can be neglected compared with 1; and since a_1 is assumed to be very large, the $+1$ term in the last row can be neglected compared with ja_1 . Only a few points are shown in the table because the general shape of the curve is evident. This portion of the L_1 contour is sketched in Fig. 7-7-5b with the mapping of the points ja_2 , $j0.5$, $j1$, and ja_1 also shown.

The negative $j\omega$ -axis maps from the same function as given by Eq. 7-7-5; but to emphasize the fact that ω is negative, this equation is rewritten as

$$G(s) \Big|_{s=-j\omega} = \frac{1}{(-j\omega)(-j\omega + 1)} \quad (7-7-6)$$

TABLE 7-7-1

ω	$\frac{1}{j\omega}$	$\frac{1}{j\omega + 1}$	$G(j\omega) = \frac{1}{j\omega} \left(\frac{1}{j\omega + 1} \right)$
a_2	$\frac{1}{a_2} \angle -90^\circ$	$\frac{1}{ja_2 + 1} = 1$	$\frac{1}{a_2} \angle -90^\circ$
0.5	$2 \angle -90^\circ$	$\frac{1}{j0.5 + 1} = 0.9 \angle -26.6^\circ$	$1.8 \angle -116.6^\circ$
1	$1 \angle -90^\circ$	$\frac{1}{j1 + 1} = 0.707 \angle -45^\circ$	$0.707 \angle -135^\circ$
a_1	$\frac{1}{a_1} \angle -90^\circ$	$\frac{1}{ja_1 + 1} = \frac{1}{a_1} \angle -90^\circ$	$\frac{1}{a_1^2} \angle -180^\circ$

For each value of ω , Eq. 7-7-6 is the conjugate of Eq. 7-7-5; therefore, Table 7-7-1 suffices for this mapping by simply changing the signs on the angles of $G(j\omega)$. The mapping of the negative $j\omega$ -axis is shown in Fig. 7-7-5b.

If the contour l_1 in the s -plane did not avoid the pole at the origin, the contour L_1 in the $[G(s)]$ -plane would go off to and return from infinity with the question as to how the contour closes left unanswered. Therefore to avoid this uncertainty, the point a moves up the $j\omega$ -axis then along a semi-circle in the right half s -plane. On this semi-circle s can be expressed as

$$s = a_2 \varepsilon^{ja_2} \quad (7-7-7)$$

where a_2 takes on values from -90° to 0° to $+90^\circ$. Since a_2 is arbitrarily small, the behavior of $G(s)$ as s approaches zero is the only concern as to how this portion of L_1 maps. For this example, $G(s)$ behaves as

$$G(s) \bigg|_{s=a_2 \varepsilon^{ja_2}} = \frac{1}{s(s+1)} \bigg|_{s=a_2 \varepsilon^{ja_2}} \approx \frac{1}{a_2} \varepsilon^{-ja_2} \quad (7-7-8)$$

Since a_2 can be made as small as desired, the term $a_2 \varepsilon^{ja_2}$ can be neglected when compared to unity. This portion of l_1 maps into a semi-circle of radius $1/a_2$ and an angle that is the negative of a_2 as shown in Fig. 7-7-5b.

The only portion of the l_1 contour not explored is the large semi-circle along which s can be expressed as

$$s = a_1 \varepsilon^{+ja_1} \quad (7-7-9)$$

where a_1 takes on values from $+90^\circ$ to 0° to -90° . Since a_1 is arbitrarily large, the behavior of $G(s)$ as s approaches infinity is the only concern as to how this portion of L_1 maps. For this example, $G(s)$ behaves as

$$G(s) \bigg|_{s=a_1 \varepsilon^{+ja_1}} = \frac{1}{s(s+1)} \bigg|_{s=a_1 \varepsilon^{+ja_1}} \approx \left(\frac{1}{a_1} \right)^2 \varepsilon^{-j2a_1} \quad (7-7-10)$$

Since a_1 can be made as large as desired, the unity term can be neglected when compared with $a_1 \varepsilon^{ja_1}$. This portion of l_1 maps into a circle of radius $(1/a_1)^2$ and an

angle that is the negative of a_1 but of twice the magnitude. The region near the origin in the $[G(s)]$ -plane of Fig. 7-7-5b is amplified and shown in Fig. 7-7-5c. As a_1 becomes larger and larger, the radius $(1/a_1)^2$ becomes smaller and smaller. This portion of the contour in the $[G(s)]$ -plane shrinks to a point as a_1 approaches infinity.

The l_1 contour in the s -plane is traversed in the clockwise direction and the arrows are added in Fig. 7-7-5a to indicate this. The arrows are also added in Fig. 7-7-5b and c to indicate the direction of traversal of L_1 .

After the L_1 contour is completed, the question of whether the closed-loop system is stable can be answered. An examination of the $G(s)$ of Eq. 7-7-4 indicates that P is equal to zero. An inspection of Fig. 7-7-5b shows that N is also zero. Eq. 7-7-3 is solved for Z as

$$Z = P - N = 0 - 0 = 0 \quad (7-7-11)$$

Since Z is zero, the closed-loop system is stable.

In this particular example, there is no difficulty in determining that N is equal to zero. In complicated examples, the problem of determining N is more involved. One convenient method to follow is to visualize a phasor originating at -1 and the arrowhead terminating on the L_1 contour. As the arrowhead traverses the L_1 contour, the net number of counterclockwise rotations of the phasor is equal to N . An even more convenient method is to draw a test line from the minus one point to infinity and to count the net number of crossings the L_1 contour makes with this line. Two such lines are drawn in Fig. 7-7-5b. Test line no. 1 does not cross the L_1 contour. The first point at which test line no. 2 crosses L_1 , the L_1 contour is going from right to left while the second point at which they cross, the L_1 contour is going from left to right. Both test lines 1 and 2 yield that $N = 0$.

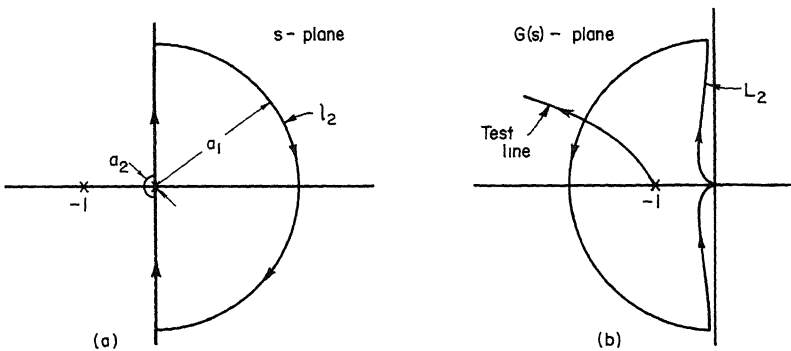


Fig. 7-7-6.

Example 7-7-2. The same $G(s)$ and $H(s)$ of Ex. 7-7-1 are used again

$$G(s) = \frac{1}{s(s+1)} \quad ; \quad H(s) = 1 \quad (7-7-12)$$

except this time the contour in the s -plane is chosen as l_2 as shown in Fig. 7-7-6a. The l_2 contour is similar to l_1 except the pole at the origin is avoided by having l_2 move around a small semi-circle in the left half s -plane. The L_2 contour is constructed in a manner similar to that described for L_1 and is shown in Fig. 7-7-6b.

A test line is drawn from -1 out to infinity. The L_2 contour crosses this line one time in such a direction that it can be deduced that $N = 1$. Since the contour in the s -plane encloses the origin, $P = 1$. The equation for Z gives

$$Z = P - N = 1 - 1 = 0 \quad (7-7-13)$$

again yielding the result that the closed-loop system is stable.

It is interesting to compare l_2 and L_2 from the standpoint of conformal mapping. Point a traverses the l_2 contour while point A (the mapping of point a) traverses the L_2 contour. Point a is arbitrarily assumed to start at $-j\infty$ thereby causing point A to start at the origin. Point a moves up the $j\omega$ -axis and point A moves along the L_2 contour. When point a reaches $-ja_2$, the l_2 path turns through 90° to the left; correspondingly, point A reaches the large semi-circle and the L_2 path also turns through 90° to the left. This is true because the mapping is conformal and the magnitude and sense of the angle must be maintained. This same type of reasoning can be repeated for the remainder of the mapping and for all other situations in which the mapping is conformal.

Example 7-7-3. The same $G(s)$ and $H(s)$ of Exs. 7-7-1 and 7-7-2 are used again:

$$G(s) = \frac{1}{s(s+1)} \quad ; \quad H(s) = 1 \quad (7-7-14)$$

In the previous two examples l_1 and l_2 enclosed the right half s -plane; and since $Z = 0$, the characteristic polynomial has no zeros in the right half s -plane and the closed-loop system is absolutely stable.

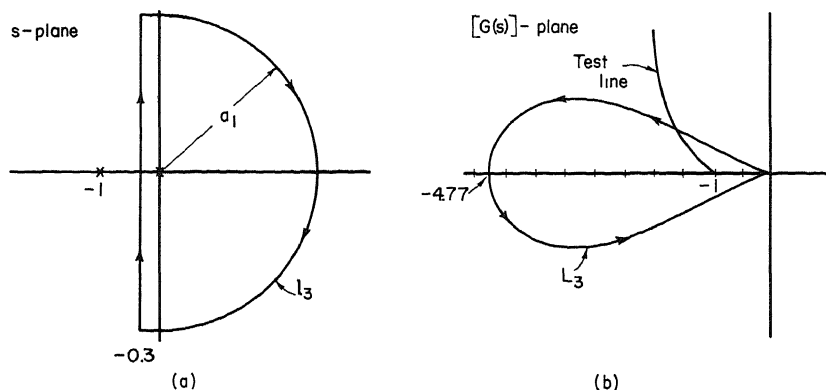


Fig. 7-7-7.

The l contour can enclose any other portion of the s -plane. In the first part of this example the l contour is shown in Fig. 7-7-7a. By making a_1 larger and larger, not only is the entire right half s -plane enclosed in the l_3 contour but also the vertical strip between $s = 0$ and $s = -0.3$.

The mapping of the vertical straight line portion of l_3 is obtained by substituting $s = -0.3 + j\omega$ in the $G(s)$ of Eq. 7-7-14.

$$\begin{aligned} G(-0.3 + j\omega) &= \frac{1}{(-0.3 + j\omega)(-0.3 + j\omega + 1)} \\ &= \frac{1}{(-0.3 + j\omega)(0.7 + j\omega)} \end{aligned} \quad (7-7-15)$$

TABLE 7-7-2

ω	$-0.3 + j\omega$	$0.7 + j\omega$	$G(0.3 + j\omega)$
0	-0.3	0.7	$\frac{1}{(-0.3)(0.7)} = -4.77$
0.5	$\frac{-0.3 + j0.5}{0.583 \angle 121^\circ}$	$\frac{0.7 + j0.5}{0.856 \angle 35.6^\circ}$	$\frac{1}{0.583 \angle 121^\circ \cdot 0.856 \angle 35.6^\circ}$ $= 2.0 \angle -156^\circ$
1.0	$\frac{-0.3 + j1}{1.05 \angle 106.7^\circ}$	$\frac{0.7 + j1}{1.22 \angle 55^\circ}$	$\frac{1}{1.05 \angle 106.7^\circ \cdot 1.22 \angle 55^\circ}$ $= 0.783 \angle -161.7^\circ$
∞	$\frac{-0.3 + j\infty}{= \infty \angle 90^\circ}$	$\frac{0.7 + j\infty}{= \infty \angle 90^\circ}$	$\frac{1}{\infty \angle 90^\circ \cdot \infty \angle 90^\circ} = 0 \angle -180^\circ$

Table 7-7-2 is formed for positive values of ω as an aid in developing $G(-0.3 + j\omega)$. These points are plotted in Fig. 7-7-7b as a portion of the L_3 contour. The portion of the line $s = -0.3 + j\omega$ for negative values of ω maps as the conjugate of the points for positive ω thus completing the L_3 contour.

A test line is drawn from -1 out to infinity. The L_3 contour crosses this line in such a direction that it can be deduced that $N = 1$. The contour in the s -plane encloses the pole at the origin in $G(s)$ and therefore $P = 1$. The equation for Z yields

$$Z = P - N = 1 - 1 = 0 \tag{7-7-16}$$

Therefore all the zeros of the characteristic polynomial are to the left of $s = -0.3$.

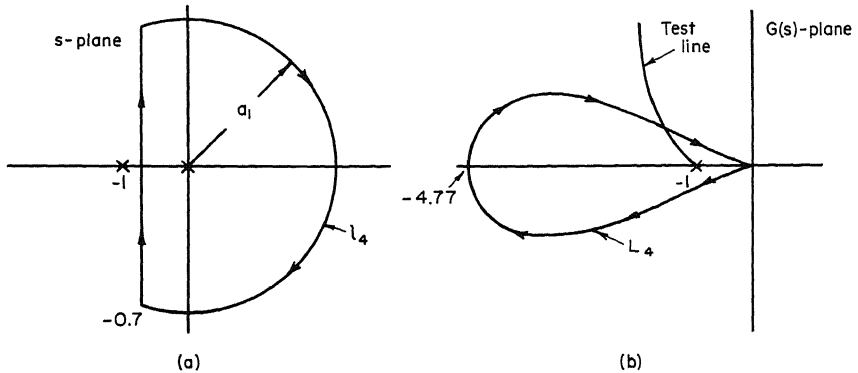


Fig. 7-7-8.

The example is continued by choosing the l contour as shown in Fig. 7-7-8a. By making a_1 larger and larger not only is the entire right half s -plane enclosed in the

l_4 contour but also is the vertical strip between $s = 0$ and $s = -0.7$. The mapping of the line $s = -0.7 + j\omega$ is obtained from the $G(s)$ of Eq. 7-7-1 as

$$\begin{aligned} G(-0.7 + j\omega) &= \frac{1}{(-0.7 + j\omega)(-0.7 + j\omega + 1)} \\ &= \frac{1}{(-0.7 + j\omega)(0.3 + j\omega)} \end{aligned} \quad (7-7-17)$$

The L_4 contour of Fig. 7-7-8b is plotted from this equation.

A test line is drawn from -1 out to infinity. The L_4 contour crosses this line in such a direction that it can be deduced that $N = -1$. Since the contour in the s -plane encloses the pole at the origin in $G(s)$, $P = 1$. The equation for Z yields

$$Z = P - N = 1 + 1 = 2 \quad (7-7-18)$$

Therefore, it can be reasoned that the characteristic polynomial has two zeros between $s = -0.3$ and $s = -0.7$.

This example can be continued by changing the l contour to subdivide the strip between $s = -0.3$ and $s = -0.7$. When $s = -0.5 + j\omega$ is used, the L contour moves out along the negative real axis and back. This is the borderline case between the L contour circling the -1 in the counterclockwise direction or in the clockwise direction. It can finally be reasoned that the characteristic polynomial has two zeros located on the line $s = -0.5 + j\omega$.

TABLE 7-7-3

ω	$G(j\omega) = \frac{(j\omega + 4)}{j\omega(j\omega - 1)}$
0	$\frac{(j0 + 4)}{j0(j0 - 1)} = j\infty$
0.5	$\frac{(j0.5 + 4)}{j0.5(j0.5 - 1)} = 7.2 \angle -236^\circ$
1	$\frac{(j1 + 4)}{j1(j1 - 1)} = 2.9 \angle -211^\circ$
2	$\frac{(j2 + 4)}{j2(j2 - 1)} = 1 \angle -180^\circ$
4	$\frac{(j4 + 4)}{j4(j4 - 1)} = 0.343 \angle -145^\circ$
10	$\frac{(j10 + 4)}{j10(j10 - 1)} = 0.1 \angle -117.5^\circ$
∞	$\frac{(j\infty + 4)}{j\infty(j\infty - 1)} = 0 \angle -90^\circ$

Example 7-7-4. The $G(s)$ and $H(s)$ of this example are

$$G(s) = \frac{K(s+4)}{s(s-1)} \quad ; \quad H(s) = 1 \quad (7-7-19)$$

This example explores absolute stability of the closed-loop system as the gain K of the open-loop system is varied. To do this, K is initially given a value of unity and Table 7-7-3 is formed.

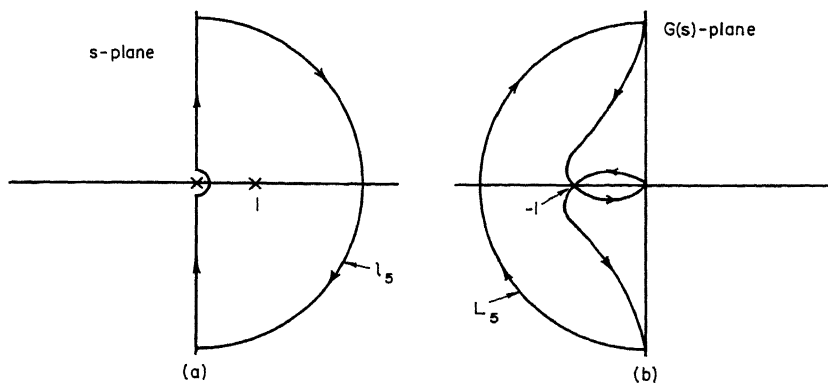


Fig. 7-7-9.

The l_s contour in the s -plane is shown in Fig. 7-7-9a and the L_s contour in the $G(s)$ plane is shown in Fig. 7-7-9b. The l_s contour encloses the pole at $s = 1$ and therefore $P = 1$. The L_s contour passes through the -1 point. If K were increased above unity, the -1 point would be inside the smaller loop with the result that $N = 1$. Under these conditions, the equation for Z yields

$$Z = P - N = 1 - 1 = 0 \quad (7-7-20)$$

which indicates that the closed-loop system is stable.

If K were decreased below unity, the -1 point would be inside the larger loop with the result that $N = -1$. Under these conditions, the equation for Z becomes

$$Z = P - N = 1 + 1 = 2 \quad (7-7-21)$$

which indicates that the closed-loop system is unstable.

A review of this example is interesting. Since $G(s)$ has a pole in the right half s -plane, the open-loop system is unstable. If the loop is closed with a gain K less than unity, the closed-loop is also unstable. However, if K is increased above unity the closed-loop system becomes stable.

This example is continued in Prob. 7-11 in which the pole at the origin in $G(s)$ is avoided by a small semi-circle in the left half s -plane.

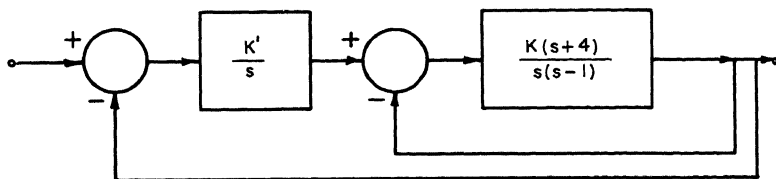


Fig. 7-7-10.

Example 7-7-5. The block diagram of Fig. 7-7-10 is considered in this example. The inner loop is the system considered in the previous example. The open-loop system is unstable but the closed-loop system is stable if K is larger than unity. K is arbitrarily chosen as 2 in this example and the problem is to determine K' such that the entire system is stable. The transfer function of the inner closed-loop system is

$$\frac{G(s)}{1 + G(s)} = \frac{2s + 8}{s^2 + s + 8} \quad (7-7-22)$$

If the outer loop is opened, the transfer function of the open-loop system is

$$G'(s) = \frac{K'(2s + 8)}{s(s^2 + s + 8)} \quad (7-7-23)$$

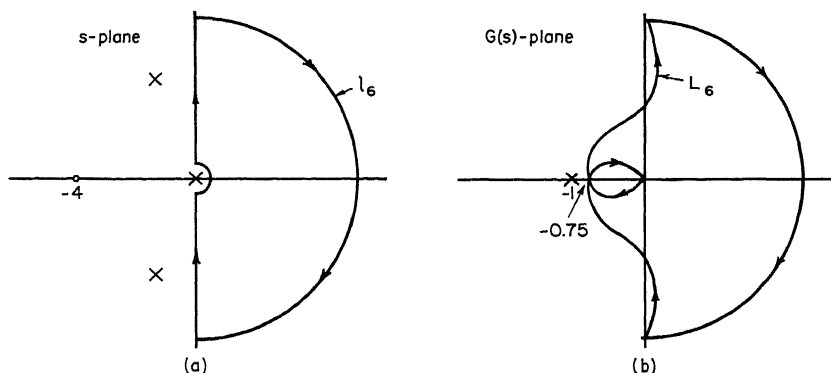


Fig. 7-7-11.

In order to plot a Nyquist diagram, K' is assigned the value of unity. The l_6 contour in the s -plane is shown in Fig. 7-7-11a and it can be seen that $P = 0$. For $K' = 1$, the L_6 contour is shown in Fig. 7-7-11b. It can be seen that $N = 0$ and from the equation for Z ,

$$Z = P - N = 0 - 0 = 0 \quad (7-7-24)$$

that the closed-loop system is stable. As K' is increased, the L_6 contour moves closer and closer to the -1 point and when K' is larger than 1.33 the -1 point is inside the L_6 contour resulting in a value of $N = -2$ and a Z of

$$Z = P - N = 0 + 2 = 2 \quad (7-7-25)$$

This example is continued as Probs. 7-12, 7-13 and 7-14 in which various values of K and K' are considered.

7-8. System Type Numbers and the Nyquist Diagrams. This article discusses the effect of the system type number on the Nyquist diagram. In this discussion it is assumed that the system has unity feedback and $G(s)$ is of the form

$$G(s) = \frac{K(s + z_1)(s + z_2) \cdots (s + z_k)}{s^N(s + p_1)(s + p_2) \cdots (s + p_j)} \quad (7-8-1)$$

where N is the system type number.

TYPE 0 SYSTEM.

Example 7-8-1. The simplest Type 0 system is given by

$$G(s) = \frac{K}{(s + p_1)} \quad (7-8-2)$$

Instead of forming a table for $G(j\omega)$ as ω is varied, the following more qualitative approach is used. The contour in the s -plane is shown in Fig. 7-8-1a. The ω is

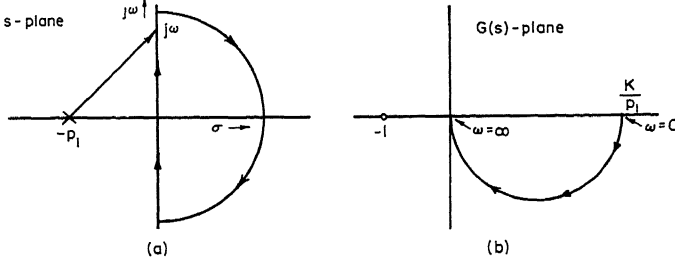


Fig. 7-8-1.

varied from 0 to $+\infty$ and the phasor from the pole at $-p_1$ to $j\omega$ is observed. When $\omega = 0$, this phasor has a value of p_1 and the plot in the $G(s)$ plane starts at K/p_1 . As ω increases, the length of the phasor increases and the length approaches ∞ as ω approaches ∞ . The angle on the phasor is zero when ω is zero and approaches $+90^\circ$ as ω approaches ∞ . Since the phasor is in the denominator of $G(s)$, the angle of $G(j\omega)$ goes from 0 to -90° .

The effect of completing the l contour in the s -plane can be visualized without actually drawing all of the L contour. From this it can be seen that the closed-loop system is absolutely stable regardless of the magnitude of K .

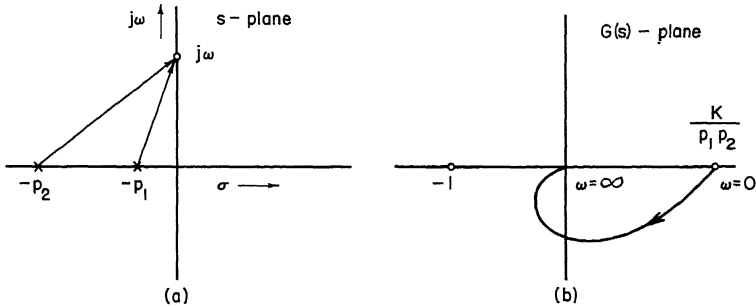


Fig. 7-8-2.

Example 7-8-2. The second Type 0 system to be explored is given by

$$G(s) = \frac{K}{(s + p_1)(s + p_2)} \quad (7-8-3)$$

The swing of the phasors in the s -plane of Fig. 7-8-2a is observed to yield the L contour in Fig. 7-8-2b. Again the effect of completing the l contour in the s -plane can be visualized without actually drawing all of the L contour. From this it can be seen that the closed-loop system is absolutely stable regardless of the magnitude of K . For a given p_1 and p_2 , however, the L contour comes closer and closer to the -1 point as K is increased and it can be intuitively seen that the relative stability decreases as K increases. This matter of relative stability is discussed in Art. 7-10.

Example 7-8-3. The last Type 0 system to be explored is given by

$$G(s) = \frac{K}{(s + p_1)(s + p_2)(s + p_3)} \quad (7-8-4)$$

The swing of the phasors in the s -plane of Fig. 7-8-3a is observed to yield the L

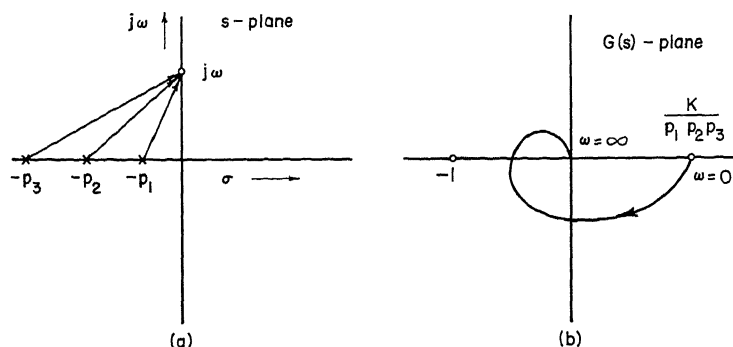


Fig. 7-8-3.

contour in Fig. 7-8-3b. As K is increased, the L contour moves until it crosses the -1 point. The closed-loop system is unstable for any K equal to or larger than this value.

In reviewing the last three examples, it should be noted that the portion of the Nyquist plot characteristic of the Type 0 system is the origination of the plot on the positive real axis for $s \rightarrow 0$. The behavior of $G(s)$ as $s \rightarrow \infty$ is determined by the number of excess poles over zeros of $G(s)$. In Ex. 7-8-1 the excess is one and Nyquist plot approaches the origin along the -90° axis as $s \rightarrow \infty$. In Ex. 7-8-2, the excess of poles over zeros in $G(s)$ is equal to 2 and the Nyquist plot approaches the origin along the negative real axis as $s \rightarrow \infty$. In Ex. 7-8-3, the excess is three and the Nyquist plot approaches the origin along the $+90^\circ$ axis.

TYPE 1 SYSTEM.

Example 7-8-4. If an integration can be added inside the loop of a Type 0 system it becomes a Type 1 system. As the first example, an s is added as a factor in the denominator of Eq. 7-8-2, resulting in

$$G(s) = \frac{K}{s(s + p_1)} \quad (7-8-5)$$

The phasors in the s -plane are shown in Fig. 7-8-4a. For small values of ω , the phasor from $-p_1$ has an angle that is essentially equal to zero; however, the angle from the pole at the origin is always equal to 90° . Therefore, for small values of ω , the $G(j\omega)$ plot starts at -90° . As ω approaches infinity, the angle on the phasor from $-p_1$ is also 90° and the $G(j\omega)$ plot approaches the origin along the negative real axis. From the Nyquist diagram it is observed that the closed-loop system is absolutely stable for all values of K .

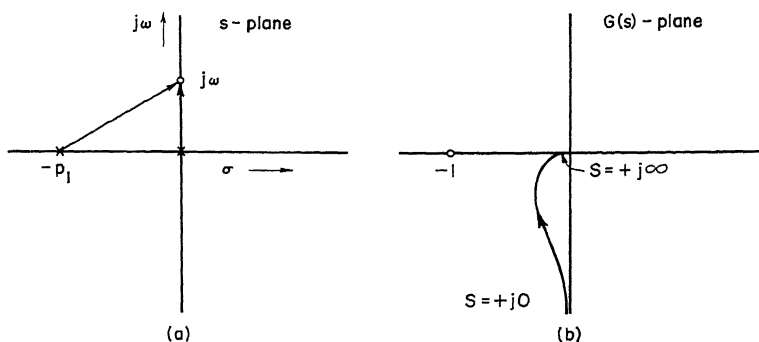


Fig. 7-8-4.

Example 7-8-5. As the second example an s is added as a factor to the denominator of Eq. 7-8-3 resulting in

$$G(s) = \frac{K}{s(s + p_1)(s + p_2)} \quad (7-8-6)$$

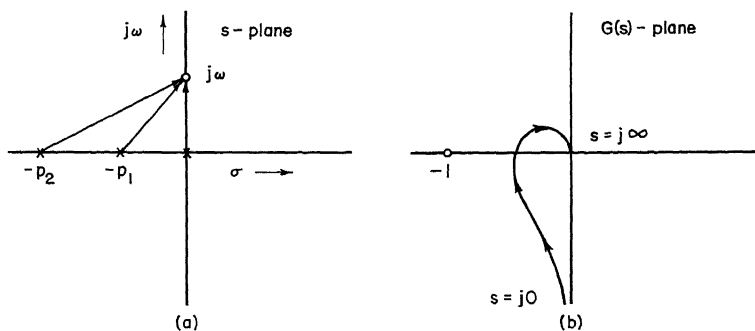


Fig. 7-8-5.

The phasors in the s -plane are shown in Fig. 7-8-5a. From the Nyquist diagram of Fig. 7-8-5b, it is observed that the closed-loop system is stable for small values of K but unstable for large values of K .

In reviewing the last two examples, it should be noted that the portion of the Nyquist plot characteristic of the Type 1 system is the origination of the plot at $-j\infty$ in the $G(s)$ plane for $s = 0$. As with the Type 0 system, the behavior of $G(s)$ as $s \rightarrow \infty$ is determined by the number of excess poles over zeros of $G(s)$.

TYPE 2 SYSTEM.

Example 7-8-6. If an integration can be added inside the loop of a Type 1 system, the result is a Type 2 system. As the first example the denominator of Eq. 7-8-5 is multiplied by s resulting in

$$G(s) = \frac{K}{s^2(s + p_1)} \quad (7-8-7)$$

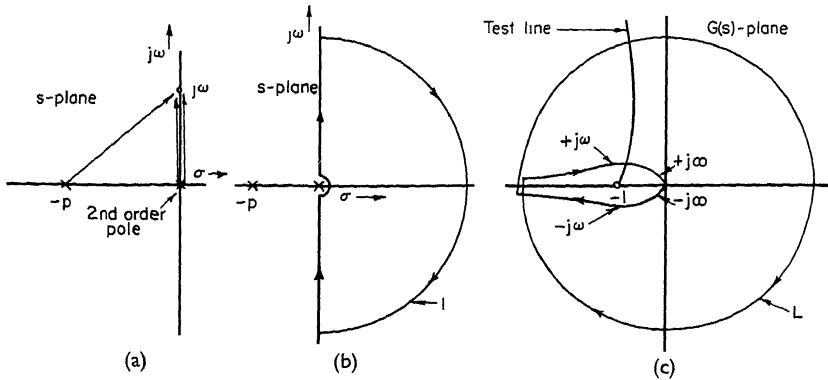


Fig. 7-8-6.

The phasors in the s -plane are shown in Fig. 7-8-6a. It is convenient to visualize two phasors originating from the 2nd-order pole at the origin thus contributing an angle of -180° to the $G(j\omega)$ plot. As ω increases, the angle on the phasor from $-p_1$ increases and the $G(j\omega)$ plot swings in the negative direction beyond -180° . The entire l contour in the s -plane is shown in Fig. 7-8-6b and the L contour in Fig. 7-8-6c. The L contour crosses the test line twice in such a direction that $N = -2$. The equation for Z yields $Z = P - N = +2$.

The 2nd-order pole at the origin contributes a phase angle of -180° to the $G(j\omega)$ plot. This fact must be accepted if a Type 2 system is required. In addition to the 2nd-order pole at the origin, any other pole (or poles) added to $G(s)$ causes the $G(j\omega)$ plot to swing in the lagging direction and produces an unstable system. To produce a stable system, a zero (or zeros) must be added so that the $G(j\omega)$ plot swings in the leading direction in the vicinity of the -1 point. This is explored in the next example.

Example 7-8-7. A zero must be added inside the pole of Eq. 7-8-7 for the system of the last example to become stable. The $G(s)$ is given by

$$G(s) = \frac{K(s + z_1)}{s^2(s + p_1)} \quad ; \quad z_1 < p_1 \quad (7-8-8)$$

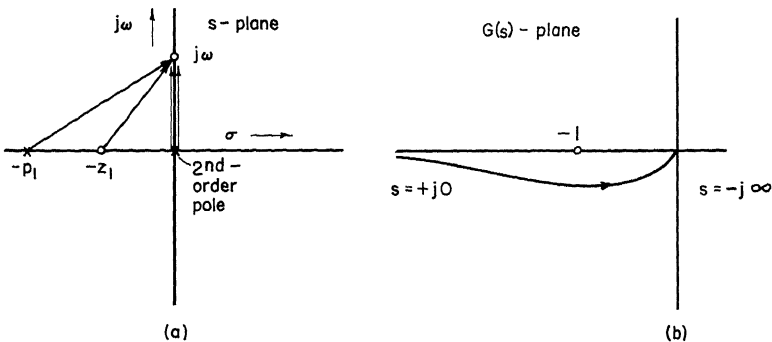


Fig. 7-8-7.

The phasors in the s -plane are shown in Fig. 7-8-7a. As s moves up the $j\omega$ -axis, the phasor from the zero swings by a larger angle than does the phasor from the pole

until $\omega \rightarrow \infty$, at which time both angles approach 90° . The corresponding Nyquist plot is shown in Fig. 7-8-7b and it can be seen that the closed-loop system is stable.

Example 7-8-8. This example considers the situation in which $G(s)$ has an excess of three poles over zeros. The $G(s)$ is given by

$$G(s) = \frac{K(s + z_1)}{s^2(s + p_1)(s + p_2)} \quad (7-8-9)$$

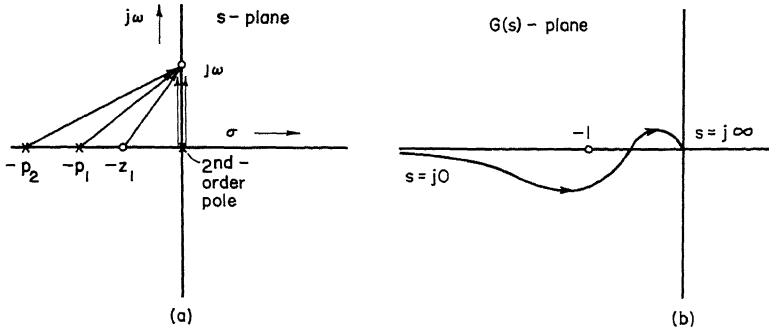


Fig. 7-8-8.

The phasors in the s -plane are shown in Fig. 7-8-8a. As s moves up the $j\omega$ -axis, the phasor from the zero must swing with a large enough angle to overcome the sum of the two angles from the poles at $-p_1$ and $-p_2$ when the $G(j\omega)$ plot is in the vicinity of the -1 point. Such a $G(j\omega)$ plot representing a stable system is shown in Fig. 7-8-8b.

Example 7-8-9. As can be anticipated, the problem of avoiding the -1 point becomes progressively more difficult as the type number of the system increases. A $G(s)$ is chosen for this example with an excess of three poles over zeros as given by

$$G(s) = \frac{K(s + z_1)(s + z_2)}{s^3(s + p_1)(s + p_2)} \quad ; \quad z_1 < z_2 < p_1 < p_2 \quad (7-8-10)$$

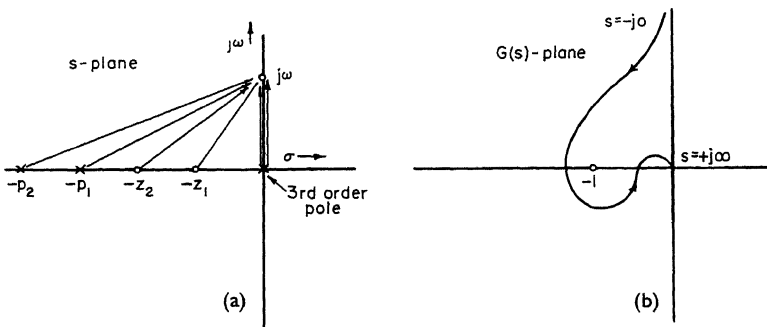


Fig. 7-8-9.

The phasors in the s -plane are shown in Fig. 7-8-9a. As s moves up the $j\omega$ -axis, the phasors from the two zeros must swing so that the sum of their two angles exceeds the sum of the two angles from the poles $-p_1$ and $-p_2$ by at least 90° when the $G(j\omega)$ plot is in the vicinity of the -1 point. Such a $G(j\omega)$ plot representing a stable system is shown in Fig. 7-8-9b. If the gain is either increased or decreased sufficiently, the closed-loop system becomes unstable.

7-9. An Additional Comment Concerning the Nyquist Diagram.

Several of the Nyquist diagrams are not completed in the previous article but only the mappings of the $+j\omega$ -axis are shown. As long as the entire Nyquist diagram is visualized, this partial mapping yields satisfactory results. To emphasize the need of visualizing the entire diagram, a Type 1 system and a Type 5 system are compared

$$G_1(s) = \frac{K}{s(s + p_1)} \quad (7-9-1)$$

and

$$G_5(s) = \frac{K}{s^5(s + p_1)} \quad (7-9-2)$$

For the Type 1 and Type 5 systems, the mapping of the $j\omega$ -axis becomes

$$G_1(j\omega) = \frac{K}{j\omega(j\omega + p_1)} \quad (7-9-3)$$

and

$$G_5(j\omega) = \frac{K}{j\omega^5(j\omega + p_1)} \quad (7-9-4)$$

respectively.

The mapping of the $j\omega$ -axis is very similar in both cases; however, the entire Nyquist diagrams are very different as shown in Fig. 7-9-1. The

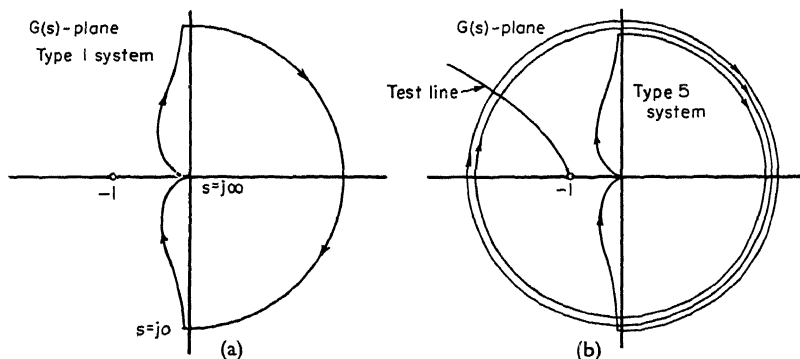


Fig. 7-9-1.

contour in the s -plane avoids the pole (or poles) at the origin by detouring around a small semi-circle in the right half s -plane similar to Fig. 7-7-5a. The plot for the Type 1 system is shown in Fig. 7-9-1a and it can be seen that the closed-loop system is stable. The plot of the Type 5 system is in Fig. 7-9-1b and the Nyquist diagram crosses the test line such that $N = -2$ and $Z = P - N = 2$ thus indicating an unstable system.

7-10. Relative Stability from the Nyquist Diagram. As soon as it has been shown that the closed-loop system is stable, the Nyquist diagram can be used to yield information concerning relative stability.

Fig. 7-10-1 indicates the Nyquist diagram for three systems. Systems no. 1 and 2 are stable. (The remainder of the diagrams are not shown but it is assumed that the completed diagrams indicate the systems are stable.) System no. 3 is unstable. The only difference among the three systems is a matter of the gain of the open-loop system. System no. 2 has a higher gain than system no. 1 and system no. 3 has a higher gain than system no. 2. When the Nyquist diagram is to the left of the -1 point, the closed-loop system is unstable. When the diagram is to the right of the -1 point, the closed-loop system is stable; however, it is intuitively evident that the closer the diagram is to the -1 point the lower the relative stability. Therefore, system no. 1 has greater relative stability than does system no. 2. This concept is developed in a more quantitative manner as the discussion proceeds.

The significance of another term that is commonly used can be understood by again referring to Fig. 7-10-1. This term is "phase margin." Instead of centering interest only on the -1 point, the unit circle is also added to the figure. Lines are drawn from the origin to the point at which the unit circle crosses the appropriate $G(j\omega)$ plot and the angle from the negative real axis in a counterclockwise direction to each such line is the phase margin. On Fig. 7-10-1, the phase margin for system no. 1 is the angle γ_1 which is positive. The phase margin for system no. 2 is the angle γ_2 which is also positive; however, the phase margin for system no. 3 is negative. For such a system to be stable, the phase margin must be positive. In a very rough sort of way, the magnitude of the phase margin of a stable system is sometimes used as an indicator of relative stability. This indicator must be used with caution since it is easy to visualize a $G(j\omega)$ plot that is close to the -1 point at some range of frequencies and yet crosses the unit circle in such a way as to yield a large phase margin. However, for many systems the phase margin can be used by the designer as an indicator of relative stability. Examples of this nature are given in Chapter 9.

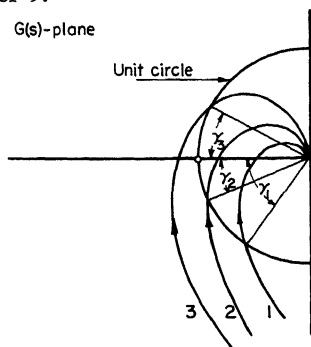


Fig. 7-10-1.

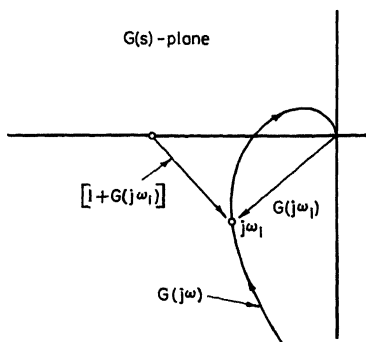


Fig. 7-10-2.

Fig. 7-10-2 indicates the Nyquist diagram for a stable system with unity feedback. The point $j\omega_1$ in the s -plane maps as $G(j\omega_1)$ in the $G(s)$ plane. The phasor from the origin to this point represents $G(j\omega_1)$. The phasor from -1 to this point represents $1 + G(j\omega_1)$. The transfer function for the closed-loop system evaluated at $s = j\omega_1$ is

$$\frac{C}{R}(j\omega_1) = \frac{G(j\omega_1)}{1 + G(j\omega_1)} \quad (7-10-1)$$

Therefore, the ratio of the two phasors of Fig. 7-10-2 as given by Eq. 7-10-1 yields information concerning the closed-loop system from the Nyquist diagram of the open-loop system. The ratio of these two phasors is a complex number. The conventional notation for the magnitude of this ratio is M and for the angle is α as given by

$$\frac{C}{R}(j\omega) = M \angle \alpha = \frac{G(j\omega)}{1 + G(j\omega)} \quad (7-10-2)$$

The locus for constant M is found in the following development. $G(j\omega)$ is a complex quantity as given by

$$G(j\omega) = U + jV \quad (7-10-3)$$

M is given by

$$M = \frac{|U + jV|}{|1 + U + jV|} \quad (7-10-4)$$

from which M^2 is found as

$$M^2 = \frac{U^2 + V^2}{(1 + U)^2 + V^2} = \frac{U^2 + V^2}{1 + 2U + U^2 + V^2} \quad (7-10-5)$$

The equation is rearranged as

$$U^2[1 - M^2] - 2M^2U - M^2 + [1 - M^2]V^2 = 0 \quad (7-10-6)$$

$$U^2 - \frac{2M^2}{1 - M^2}U - \frac{M^2}{1 - M^2} + V^2 = 0 \quad (7-10-7)$$

The term $M^2/(1 - M^2)^2$ is added to both sides of the equation and it is finally written as

$$\left(U - \frac{M^2}{1 - M^2}\right)^2 + V^2 = \left(\frac{M}{1 - M^2}\right)^2 \quad (7-10-8)$$

This equation is recognized as the equation for a circle with a radius

$$R = \left| \frac{M}{1 - M^2} \right| \quad (7-10-9)$$

and center on the real axis at

$$U = \frac{M^2}{1 - M^2} \tag{7-10-10}$$

It is a simple matter to calculate the radius and center location versus M . A few such calculations are summarized in Table 7-10-1.

TABLE 7-10-1

M	Radius = $\left \frac{M}{1 - M^2} \right $	Center = $\frac{M^2}{1 - M^2}$
0.5	0.666	+0.333
0.66	1.2	+0.802
1.0	∞	∞
1.3	1.88	-2.45
1.7	0.9	-1.53
2.0	0.666	-1.33
5.0	0.209	-1.04

A family of constant M circles is shown in Fig. 7-10-3. An inspection of Table 7-10-1 and Fig. 7-10-3 indicates that as M becomes large compared with 1, the M -circles shrink down on the -1 point. Similarly

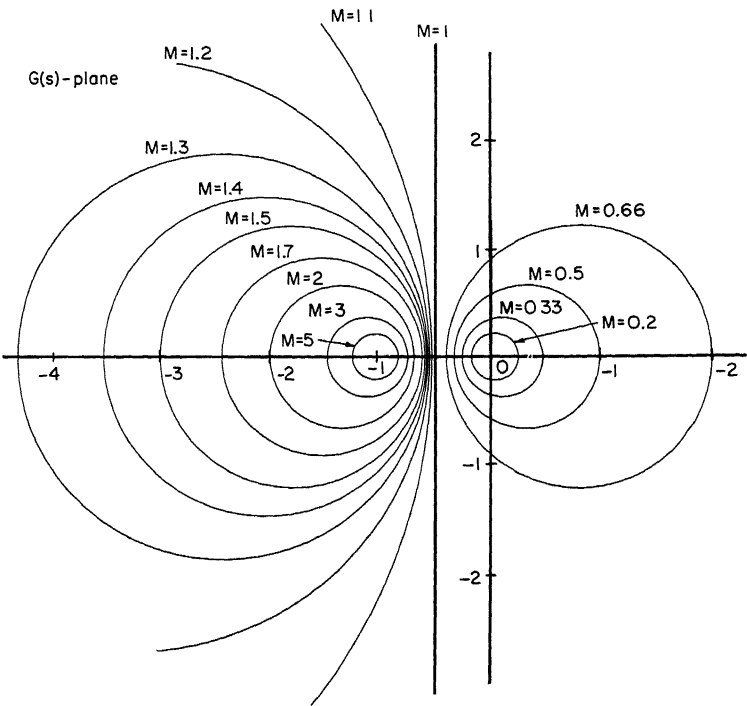


Fig. 7-10-3.

as M becomes small compared with 1, the M -circles shrink down on the origin. At first thought, it seems that the curve for $M = 1$ can not be determined from the infinite radius and center at infinity as given in Table 7-10-1. However, when it is remembered that $M = 1$ is the locus of points equidistance from the -1 point and from the origin, it can be seen that the $M = 1$ curve shown in Fig. 7-10-3 is correct.

The locus for constant a can also be determined from

$$M/\underline{a} = \frac{G(j\omega)}{1 + G(j\omega)} = \frac{U + jV}{1 + U + jV} \quad (7-10-11)$$

The angle a can be written as

$$a = \tan^{-1} \frac{V}{U} - \tan^{-1} \frac{V}{1 + U} \quad (7-10-12)$$

The $\tan a$ is defined as N

$$N = \tan a = \tan \left[\tan^{-1} \frac{V}{U} - \tan^{-1} \frac{V}{1 + U} \right] \quad (7-10-13)$$

The trigonometric identity

$$\tan (A - B) = \frac{\tan A - \tan B}{1 + \tan A \tan B} \quad (7-10-14)$$

yields for N

$$N = \frac{\frac{V}{U} - \frac{V}{1 + U}}{1 + \left(\frac{V}{U}\right)\left(\frac{V}{1 + U}\right)} \quad (7-10-15)$$

The numerator and denominator of the right side are multiplied by $U(1 + U)$

$$N = \frac{V}{U^2 + U + V^2} \quad (7-10-16)$$

and the equation is rearranged as

$$U^2 + U + V^2 - \frac{1}{N} V = 0$$

The quantity $1/4 + (1/2N)^2$ is added to both sides of the equation which is finally written as

$$\left(U + \frac{1}{2}\right)^2 + \left(V - \frac{1}{2N}\right)^2 = \left[\sqrt{\frac{1}{4} + \left(\frac{1}{2N}\right)^2}\right]^2 \quad (7-10-17)$$

This equation is recognized as the equation for a circle with a radius

$$R = \sqrt{\frac{1}{4} + \left(\frac{1}{2N}\right)^2} \quad (7-10-18)$$

and the center of the circle has the coordinates

$$U = -\frac{1}{2} ; \quad V = \frac{1}{2N} \quad (7-10-19)$$

An examination of the coordinates of the center of the circle and the radius of the circle shows each circle passes through the origin and the -1 point regardless of the value of N . Therefore, each member of the family of circles can be drawn as soon as $V = 1/2N$ is determined. A few constant N circles are shown in Fig. 7-10-4 for various values of a .

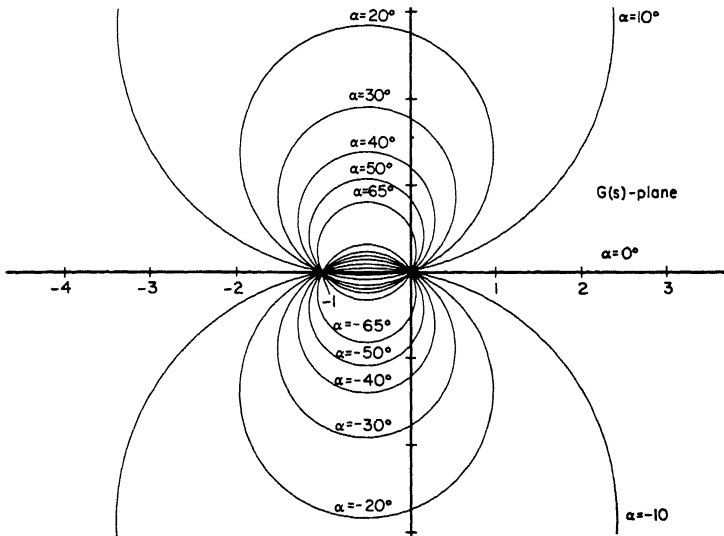


Fig. 7-10-4.

7-11. Example Using M and N Circles. When the M and N circles are added to the Nyquist diagram, information about the sinusoidal steady-state response of the closed-loop system is available by inspection. This process is demonstrated by an example

$$G(s) = \frac{1.5}{s(s+1)(s+2)} ; \quad H(s) = 1 \quad (7-11-1)$$

The plot of $G(j\omega)$ is shown in Fig. 7-11-1a and b. In Fig. 7-11-1a the M circles are added and in Fig. 7-11-1b the N circles are added. Fig. 7-11-1c shows the values of M vs. frequency that are obtained from the M circles in Fig. 7-11-1a. Fig. 7-11-1c also shows the values of phase for the closed-loop system as obtained from Fig. 7-11-1b.

It should be noted that the N circles are multivalued in that any multiple of 180° can be added to an angle and the tangent remains the

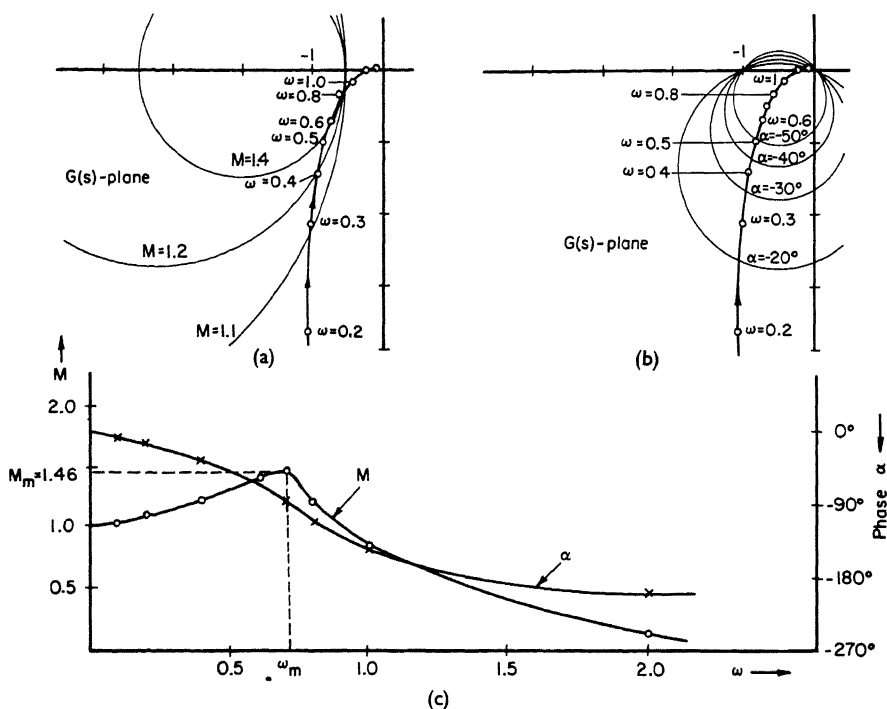


Fig. 7-11-1.

same. It can be reasoned that the phase angle should be continuous and the phase angle at $\omega = 0$ has a value of zero. As ω increases the necessary multiples of 180° can be added to (or subtracted from) the phase in order to yield a continuous curve.

7-12. Non-Unity Feedback Systems. When the transfer function of the feedback is not unity, the procedure of the previous articles must be modified. The closed-loop transfer function is written as

$$\frac{C}{R} = \left[\frac{GH}{1 + GH} \right] \frac{1}{H} \quad (7-12-1)$$

From this equation, it can be seen that C/R contains two factors. The factor inside the brackets is obtained from the M and N circles by plotting the product GH on the open-loop diagram. This factor is then multiplied by $1/H$ to obtain C/R .

7-13. Inverse Complex-plane Diagram. A transfer function is defined as the ratio of the transformed output of a system to the transformed input. Even though the transfer function is defined this way, it is sometimes more convenient to use the inverse of the transfer function rather than the transfer function itself. For example, the closed-loop transfer function is

$$\frac{C}{R} = \frac{G}{1 + GH} \quad (7-13-1)$$

The inverse of this equation is

$$\left[\frac{C}{R}\right]^{-1} = \frac{1 + GH}{G} = \frac{1}{G} + H \quad (7-13-2)$$

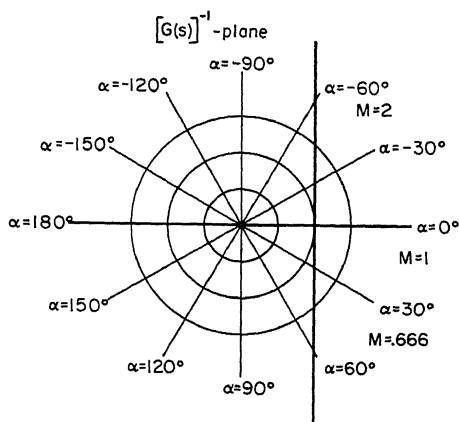


Fig. 7-13-1.

The fact that G and H appear as separate terms indicates the advantage of the inverse plot as far as graphical manipulations is concerned.

In the special case of unity feedback, Eq. 7-13-2 becomes

$$\left[\frac{C}{R}\right]^{-1} = \frac{1}{G} + 1 \quad (7-13-3)$$

The quantity $[C/R]^{-1}$ is the distance from the -1 point to any point on the $1/G(j\omega)$ plot. Therefore, the loci of constant M are circles with centers at the -1 point and with radii equal to $1/M$ as shown in Fig. 7-13-1. The loci of constant N are straight lines drawn through the -1 point as in Fig. 7-13-1.

The Nyquist criterion of stability and all the other information obtainable from the direct plots can also be obtained from the inverse plots. Since the direct plots lend themselves more readily to the understanding of the Bode plots, which present the same information in a more convenient form by using logarithms, the inverse diagrams are not featured here. However, several of these concepts are included in problems. The subject of Bode diagrams is discussed at length in Chapter 9.

PROBLEMS

7-1. In order to show that a polynomial can have all positive coefficients and yet have zeros in the right half plane, the following polynomial is built up. The polynomial has two zeros s_1 and $s_2 = +1 \pm j2$ and also contains a third zero $s = -a$. Find the range on a such that all the coefficients of the resulting cubic are positive.

7-2. Use a Routh table on each of the following polynomials as assigned and determine how many zeros each polynomial has in the right half plane, in the left half plane and on the $j\omega$ -axis.

- (a) $s^6 + 11s^5 + 47s^4 + 101s^3 + 116s^2 + 68s + 16$
- (b) $s^6 + 2s^5 + 3s^4 + 4s^3 + 5s^2 + 6s + 7$
- (c) $s^7 + s^6 + 7s^5 + 5.2s^4 + 12.25s^3 + 3.25s^2 + 4.5s + 0.45$
- (d) $s^6 + 10s^5 + 6.6s^4 + 16s^3 + 8.2s^2 + 2s + 1$
- (e) $s^6 + 26s^5 + 299s^4 + 1812s^3 + 6086s^2 + 10,780s + 7696$
- (f) $2s^7 + s^6 + 6s^5 + 3s^4 + 8s^3 + 4s^2 + 4s + 2$
- (g) $s^6 + 12s^5 + 58s^4 + 138s^3 + 161s^2 + 126s + 104$
- (h) $s^6 + 8s^5 + 38s^4 + 246s^3 + 1253s^2 + 4070s + 3016$
- (i) $s^8 + 8s^7 + 28s^6 + 58s^5 + 80s^4 + 76s^3 + 49s^2 + 20s + 4$
- (j) $s^7 + 2s^6 + 2.5s^5 + 5s^4 + 3s^3 + 6s^2 + s + 2$
- (k) $s^6 + 3s^5 + 4s^4 + 12s^3 + 10s^2 + 7s + 5$
- (l) $s^7 + 3s^6 + 4s^5 + 5s^4 + 4s^3 + 5s^2 + 3s + 2$
- (m) $s^5 + 2s^4 + 3s^3 + 6s^2 + 2s + 4$
- (n) $s^7 + s^6 + s^5 + s^4 + 2s^3 + 2s^2 + 2s + 2$
- (o) $s^5 + 2s^4 + 6s^3 + 24s^2 + 100s + 500$
- (p) $2s^6 + 3s^5 + 4s^4 + 4s^3 + 10s^2 + 15s + 20$

7-3. By methods explained in conjugation with Ex. 7-3-6, determine if each of the following polynomials has its zeros on the negative real axis.

- (a) $s^3 + 9s^2 + 26s + 24$
- (b) $s^3 + 4s^2 + 4s + 3$
- (c) $s^4 + 2s^3 + 3s^2 + 2s + 1$
- (d) $s^4 + 8s^3 + 21s^2 + 22s + 8$

7-4. Error rate and integral compensation are combined in a single system and the resulting system has the characteristic polynomial

$$Js^3 + (D + K_1)s^2 + K_2s + K_3$$

Determine the relationship among the coefficients to yield a stable system.

7-5. Polynomials (a), (e), and (i) of Prob. 7-2, are Hurwitz and hence could be the characteristic polynomial of stable control systems. It is desired to determine the real part of the zeros closest to the $j\omega$ -axis by the methods explained in Art. 7-4.

7-6. Factor polynomials (a), (e), and (i) of Prob. 7-2, by using the methods of Art. 7-4.

7-7. Given the functional relationship

$$F(s) = \frac{1}{s}$$

Map the following lines into the $F(s)$ -plane:

$$s = j\omega \quad ; \quad s = 1 + j\omega \quad ; \quad s = 2 + j\omega \quad ; \quad s = \sigma + j1 \quad \text{and} \quad s = \sigma + j2$$

7-8. Repeat Prob. 7-7 for the following functions:

- (a) $F(s) = s^2 + s$
- (b) $F(s) = \frac{1}{s+1}$
- (c) $F(s) = \frac{1}{(s+1)(s+2)}$
- (d) $F(s) = \frac{s+1}{s(s+2)}$

7-9. Develop an equation similar to Eq. 7-7-3 except make the following changes in the development leading to the equation. The contour in the s -plane is to be traversed in the counterclockwise direction and N is defined as positive for a clockwise rotation of the contour about the $-1 + j0$ point. Use the new equation to reevaluate the following: Exs. 7-7-2, 7-7-3, 7-7-4, and 7-7-5.

7-10. Repeat Prob. 7-9 except the contour in the s -plane is traversed in the clockwise direction. Use the same definition for N as in Prob. 7-9.

7-11. In Ex. 7-7-4 complete the contour in the s -plane so that the pole at the origin in $G(s)$ is avoided by a small semi-circle in the left half plane. Investigate the stability of the closed-loop system as a function of K .

7-12. For the system of Ex. 7-7-5 shown in Fig. 7-7-10, K is given a value of 3. Find the range on K' such that the entire closed-loop system is stable.

7-13. Repeat Prob. 7-12 for $K = 4$.

7-14. In Prob. 7-12, the value of K is given a value of 0.5. Investigate whether it is possible to produce a stable closed loop system as K' is varied. If so, find the range on K' .

7-15. By using a Nyquist diagram, determine the range on K such that the closed-loop system is stable. The system is defined by

$$G(s) = \frac{K}{s(s+1)(s+2)} \quad ; \quad H(s) = 1$$

Once K is determined, form the characteristic equation and check the result by using Routh's criterion.

7-16. In Prob. 7-15, choose K as one-half its maximum possible value. Reshape the contour in the s -plane by moving up lines parallel to the $j\omega$ -axis such as $s = -0.1 + j\omega$, $s = -0.2 + j\omega$, etc., instead of the $j\omega$ -axis. Locate the zeros of the characteristic polynomial of the closed-loop system within one of these vertical strips as σ is varied in 0.1 steps.

7-17. A system is defined by

$$G(s) = \frac{K(s+1.5)}{s(s+1)(s+2)} \quad ; \quad H(s) = 1$$

Investigate whether it is possible to produce an unstable closed-loop system as K is varied. If so, find the range on K .

7-18. In Prob. 7-17, let $K = 3$ locate the zeros of the characteristic polynomial within vertical strip in the s -plane as suggested in Prob. 7-16.

7-19. Repeat Prob. 7-17 for the system defined by

$$G(s) = \frac{K(s+6)}{s(s+1)(s+2)} \quad ; \quad H(s) = 1$$

7-20. Repeat Prob. 7-17 for the system defined by

$$G(s) = \frac{K(s-0.5)}{s(s+1)(s+2)} \quad ; \quad H(s) = 1$$

7-21. A system is defined by

$$G(s) = \frac{K(s+0.5)}{s(s-2)} \quad ; \quad H(s) = 1$$

Investigate whether it is possible to produce a stable closed-loop system as K is varied. If so, find the range on K .

7-22. Repeat Prob. 7-21 for the system defined by

$$G(s) = \frac{K(s + 0.5)}{s(s - 1)(s - 2)} \quad ; \quad H(s) = 1$$

7-23. Repeat Prob. 7-21 for system defined by

$$G(s) = \frac{K(s + 0.5)(s + 1.5)}{s(s - 1)(s - 2)} \quad ; \quad H(s) = 1$$

7-24. In Prob. 7-21 let $K = 4$. The resulting $G(s)$ replaces $G(s)$ in the inner loop of Fig. 7-7-10. Determine if the entire closed-loop can be made stable by varying K' . If so, find the range on K' .

7-25. Repeat Prob. 7-24 for $K = 1$.

7-26. A system is defined by

$$G(s) = \frac{(s + 1)(s + 2)}{s(s^2 + 1)} \quad ; \quad H(s) = 1$$

Determine whether the closed-loop system is stable by using the Nyquist diagram. The three poles on the $j\omega$ -axis are to be avoided by small semi-circles in the right half plane.

7-27. Repeat Prob. 7-26 but avoid the poles on the $j\omega$ -axis by small semi-circles in the left half plane.

7-28. Repeat Prob. 7-26 for the system

$$G(s) = \frac{(s + 1)}{s(s^2 + 1)(s + 10)} \quad ; \quad H(s) = 1$$

7-29. Repeat Prob. 7-26 for the system

$$G(s) = \frac{(s + 1)(s + 2)}{s(s^2 + 1)(s + 10)} \quad ; \quad H(s) = 1$$

7-30. Plot the general shape of the complete Nyquist diagram for

$$G(s) = \frac{K}{s^4(s + p_1)} \quad ; \quad H(s) = 1$$

and determine Z for the closed-loop system.

7-31. Repeat Prob. 7-30 for

$$G(s) = \frac{K}{s^7(s + p_1)} \quad ; \quad H(s) = 1$$

7-32. Repeat Prob. 7-30 for

$$G(s) = \frac{K}{s^9(s + p_1)} \quad ; \quad H(s) = 1$$

7-33. Ex. 6-14-1 concerns a system with $J = 1.445$; $D = 0.289$ and $K = 1.448$. In this example ζ is calculated to be 0.1. Plot the Nyquist diagram for this system. In order to improve the relative stability of this system, an error rate damping term is added to this system with $K_1 = 1.156$ (from Ex. 6-16-1). With K_1 added, the resulting $\zeta = 0.5$. Plot the Nyquist diagram for this system on the same sheet of paper as Ex. 6-14-1 to compare the results. To the system described by this second set of parameters, a K_2 is added. Use the relationship developed in Prob. 7-4 and set K_2 at the smallest value that produces an unstable system. Plot the Nyquist diagram of this system on the same sheet with the other two.

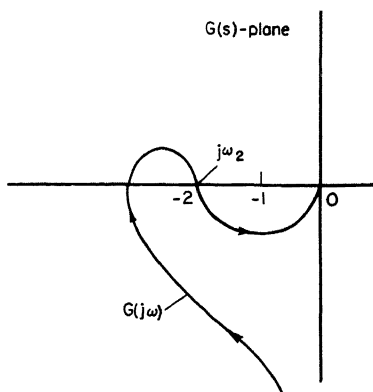


Fig. 7-P-34.

7-34. A unity feedback system has a Nyquist diagram as shown in Fig. 7-P-34. For this system, $P = 0$ and when the entire diagram is completed it is found that the closed-loop system is stable. However, this situation needs examining. The closed-loop system is opened as in Fig. 7-1-1 and a sinusoidal signal of unity amplitude with ω set at a value such that $G(j\omega_2) = -2$ is applied at E . This signal is left until steady state is reached and the signal at E' is in phase with E but has twice the amplitude. If this new signal is applied at E , then E' has an amplitude of 4, and so on. This type of reasoning indicates the closed-loop system is unstable yet from the Nyquist diagram the system is known to be stable. Explain this situation.

7-35. For the system of Prob. 7-15 with $K = 3$, add the M and N circles to the Nyquist diagram and from the resulting figure draw curves of M and α for the closed-loop system.

7-36. For the system of Prob. 7-17 with $K = 4$, add the M and N circles to the Nyquist diagram and from the resulting figure draw a curve of M and α for the closed-loop system.

7-37. Add M and N circles to the Nyquist diagrams of the systems of any of the examples and other problems as assigned.

7-38. The system of Prob. 7-15 is unstable with $K = 8$. Even though this is true, the Nyquist diagram can still be plotted, the M and N circles added, and curves of M and α vs. frequency can be plotted. What is the significance of these curves?

7-39. For the inverse complex-plane diagram, the stability criterion can be written as

$$N = P_H + Z_G - Z$$

where

P_H are the poles of $H(s)$ included in the contour.

Z_G are the zeros of $G(s)$ included in the contour.

N is the number of counterclockwise rotations of the inverse plane plot about the origin.

Z is the zeros of the characteristic polynomial included in the contour.

Develop this criterion.

7-40. Replot any of the examples or problems as assigned and determine stability from the inverse plane plot.

7-41. Add M and N circles to any of the inverse plane plots of Prob. 7-40 and obtain curves of M and α vs. frequency.

8

DESIGN USING NYQUIST DIAGRAMS

8-1. Introduction. As explained in Chapter 7, the Nyquist diagram can be used to determine absolute stability. In addition to this, if the closed-loop system is stable, the M and N circles can be added and information can be obtained regarding relative stability.

The simplest control the designer has over a system is the gain adjustment of the open-loop system. In many systems, as the gain is increased, the closed-loop system eventually becomes unstable. Examples 7-8-3 and 7-8-5 demonstrate this situation. In other systems, an increase in gain may produce a stable system, as in Ex. 7-7-4. In still other systems there is a range (or ranges) of K for which the system is stable. Example 7-8-9 is such a system.

Even when the system is stable, a variation in K produces a variation in relative stability. The system of Art. 7-11 demonstrates this point. The basic system used in this example is

$$G(s) = \frac{K}{s(s+1)(s+2)} \quad (8-1-1)$$

In Art. 7-11, K is given the value 1.5. From the Nyquist diagram of Fig. 7-11-1a it can be seen that M_m is approximately 1.4. If K is increased, M_m also increases. If K is decreased, M_m also decreases.

As discussed in Chapter 6, a "rule of thumb" is often used, to the effect that M_m should be set in the range of 1.2 to 1.4. If the designer wishes to set the gain K so that M_m is 1.3, he can simply reduce the gain until the $G(j\omega)$ plot is tangent to the $M = 1.3$ circle. Because of the nature of the $G(j\omega)$ plot, this adjustment is essentially a trial-and-error process. However, a direct method of setting the gain is discussed in the next section.

8-2. Direct Method of Setting Gain for a Given M_m . The proper gain K for a given M_m value can be determined from a graphical procedure. The following discussion refers to Fig. 8-2-1 in which M_m is assumed larger than one and only the magnitudes are shown. The radius and center of the M -circle are those determined in Chapter 7

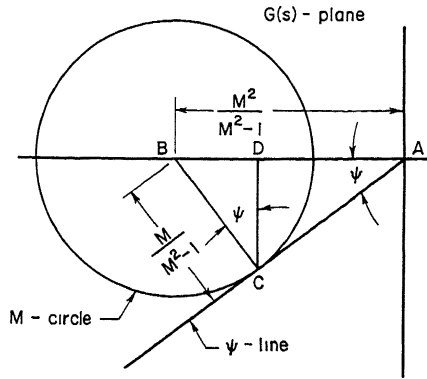


Fig. 8-2-1.

and are shown in the figure. The angle ψ is defined as the angle DAC and the $\sin \psi$ is determined as

$$\sin \psi = \frac{\frac{M}{M^2 - 1}}{\frac{M^2}{M^2 - 1}} = \frac{1}{M} \quad (8-2-1)$$

Therefore this angle depends exclusively on the value of M . From the geometry of the figure, the angle BCD also is equal to ψ . The line BD has the length

$$BD = BC \sin \psi = \frac{M}{M^2 - 1} \times \frac{1}{M} = \frac{1}{M^2 - 1} \quad (8-2-2)$$

The line AD has the length

$$AD = AB - BD = \frac{M^2}{M^2 - 1} - \frac{1}{M^2 - 1} = 1 \quad (8-2-3)$$

All this development is combined to determine the gain K of a system as explained in the following example.

Example 8-2-1. The system to be studied is

$$G(s) = \frac{K}{s(s+1)(s+2)} ; \quad H(s) = 1 \quad (8-2-4)$$

and K is to be set so that $M_m = 1.3$. The work is systematized in the following steps:

1. The value of K is set originally at 1 for convenience and the plot of $G(j\omega)$ is constructed as shown in Fig. 8-2-2a.
2. A line is drawn through the origin at an angle $\psi = \sin^{-1}\left(\frac{1}{M}\right)$.
3. A circle is constructed that has its center on the negative real axis and is tangent to both the $G(j\omega)$ plot and to the ψ line. This circle is located by trial and error.
4. A line is dropped from Point C perpendicular to the negative real axis thus locating Point D .

5. If the gain were properly adjusted, Point D would be the -1 point. The proper gain constant is calculated so that this is so. For example in Fig. 8-2-2a the length $AD = 0.77$. If the gain K is set

$$K = \frac{1}{AD} = \frac{1}{0.77} = 1.3 \quad (8-2-5)$$

then the Point D coincides with the -1 point.

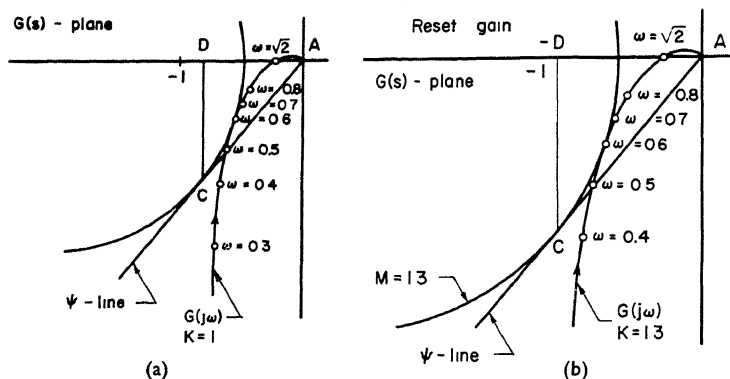


Fig. 8-2-2.

The scale of Fig. 8-2-2a is changed to adjust the figure to the K determined in Step 5. The final system with readjusted gain is shown in Fig. 8-2-2b.

8-3. System Compensation. In many situations a gain adjustment by itself is not sufficient to produce a satisfactory system. A very obvious example of this is a system that is absolutely unstable for all values of K . A specific example of this is given by

$$G(s) = \frac{K(s + z_1)}{s^2(s + p_1)} ; \quad H(s) = 1 \quad (8-3-1)$$

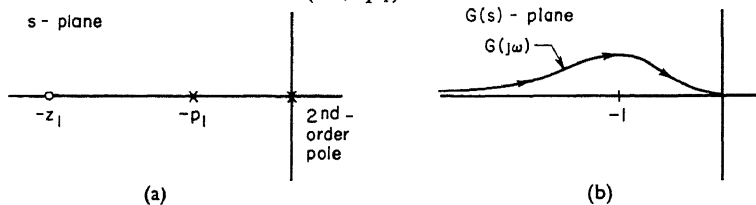


Fig. 8-3-1.

where the pole at $-p_1$ is inside the zero at $-z_1$, as shown in Fig. 8-3-1a. The $G(j\omega)$ locus is shown in Fig. 8-3-1b. In order to obtain satisfactory results from this system, the entire $G(j\omega)$ locus must be reshaped to move it to the other side of the -1 point. Such a reshaping of $G(j\omega)$ is known as compensation.

Even when it is possible to obtain a certain value of M_m by gain adjustment, the system may be unsatisfactory for other reasons. Example 8-2-1 is considered. For an $M_m = 1.3$, a $K = 1.3$ was determined, yielding

$$G(s) = \frac{1.3}{s(s+1)(s+2)} \quad (8-3-2)$$

This system has a velocity constant

$$K_v = \lim_{s \rightarrow 0} sG(s) = 0.65 \quad (8-3-3)$$

and the steady-state error with a unit-ramp function applied to the input is

$$e(t)_{ss} = \frac{1}{K_v} = \frac{1}{0.65} = 1.54 \quad (8-3-4)$$

In certain situations it may be that this steady-state error is too large and that the velocity constant is too small. Therefore, in order to improve the system, compensation is required.

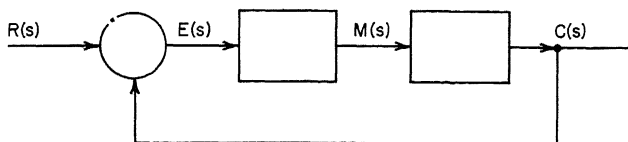


Fig. 8-3-2.

Compensation can be divided into two general classes, series compensation and feedback compensation. An example of series compensation is shown in Fig. 8-3-2. $G_2(s)$ is referred to as the fixed portion of the system. This is the portion of the system that does not yield satisfactory results before it is compensated. $G_1(s)$ is the compensation device that very often takes the form of an electric network.

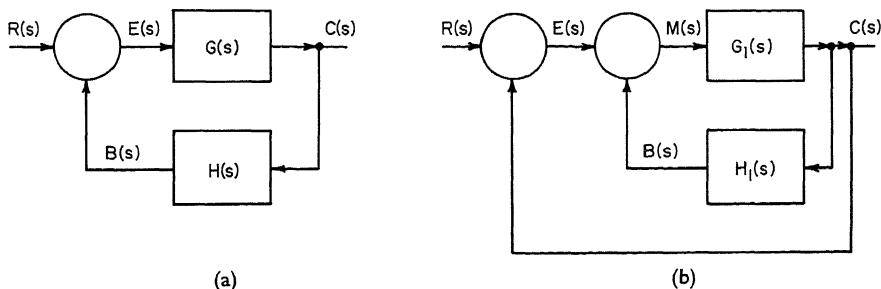


Fig. 8-3-3.

An example of feedback compensation is shown in Fig. 8-3-3a. The system with unity feedback does not yield satisfactory results, and therefore a compensation device is added to the feedback loop. Sometimes the feedback element is used to reshape the $G(j\omega)$ locus, and the system of Fig. 8-3-3a becomes a portion of some other more elaborate system as shown in Fig. 8-3-3b. Feedback compensation is often referred to as parallel compensation.

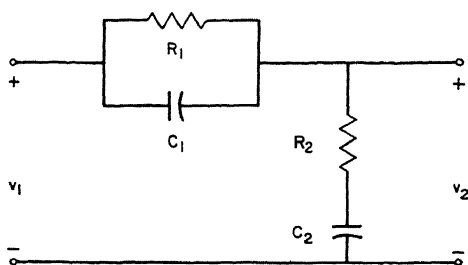


Fig. 8-4-1.

8-4. Phase Lead Network. One of the common networks used for compensation is the phase lead network shown in Fig. 8-4-1. The transfer function for this network can be written

$$G_1(s) = \frac{V_2}{V_1} = \frac{R_1 R_2 C_1 s + R_2}{R_1 R_2 C_1 s + R_1 + R_2} \quad (8-4-1)$$

or

$$G_1(s) = \frac{V_2}{V_1} = \frac{s + 1/R_1 C_1}{s + (R_1 + R_2)/R_1 R_2 C_1} \quad (8-4-2)$$

With the following definitions

$$z_1 = \frac{1}{R_1 C_1} \quad ; \quad p_1 = \frac{R_1 + R_2}{R_1 R_2 C_1} \quad (8-4-3)$$

the transfer function can be written as

$$G_1(s) = \frac{V_2}{V_1} = \frac{s + z_1}{s + p_1} \quad (8-4-4)$$

From Eqs. 8-4-3, the ratio of z_1 to p_1 is

$$\frac{z_1}{p_1} = \frac{R_2}{R_1 + R_2} \quad (8-4-5)$$

Therefore the zero is always closer to the origin than the pole and the magnitude of the ratio of z_1 to p_1 can be changed by varying R_1 and R_2 . Although there is no theoretical limit to this ratio, practical considerations do impose a limit. As can be seen from the circuit of Fig. 8-4-1 and from Eq. 8-4-4, the ratio of V_2 to V_1 at zero frequency is equal to the ratio of z_1 to p_1 . Too small a ratio means that the lower frequencies are attenuated by too large a factor. For this reason this ratio is often given a value of 1/10 and with this arbitrary choice Eq. 8-4-4 becomes

$$G_1(s) = \frac{V_2}{V_1} = \frac{s + z_1}{s + 10z_1} \quad (8-4-6)$$

In order to plot a universal curve for this, the equation is normalized by the change of variable

$$p = \frac{s}{z_1} = \frac{\sigma}{z_1} + j\frac{\omega}{z_1} = x + jy \quad (8-4-7)$$

and the equation becomes

$$G_1(p) = \frac{V_2(p)}{V_1(p)} = \frac{p + 1}{p + 10} \quad (8-4-8)$$

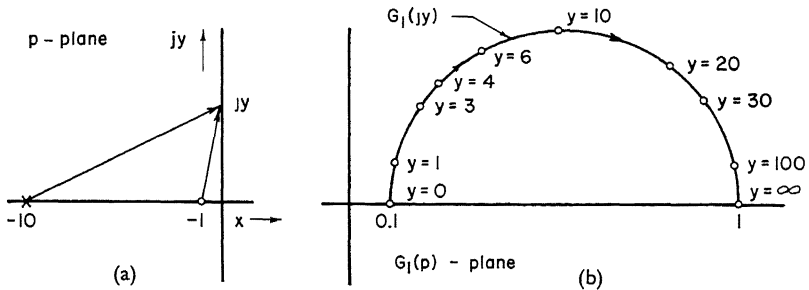


Fig. 8-4-2.

A plot of $G_1(jy)$ can be obtained from this equation. As s moves up the $j\omega$ -axis, p moves up the jy -axis as shown in Fig. 8-4-2a. Since the zero is inside the pole, the phasor in the numerator of $G_1(jy)$ swings faster than the phasor in the denominator, and the result is the leading phase characteristic of Fig. 8-4-2b. From the standpoint of complex function theory, Eq. 8-4-8 is a linear fractional transformation that maps circles into circles with straight lines as a limiting case. Therefore the curve of Fig. 8-4-2b is a semicircle.

The process of unnormalization is explained in the following example.

Example 8-4-1. In the system

$$G_2(s) = \frac{K}{s(s+1)(s+2)} \quad (8-4-9)$$

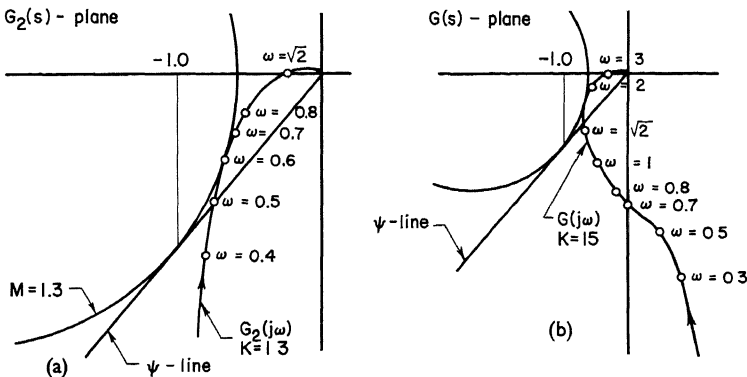


Fig. 8-4-3.

the K is set at 1.3 in Ex. 8-2-1 so that M_m is equal to 1.3. This system has a velocity constant $K_v = 0.65$. It is desired to increase the velocity constant by using a lead network and yet maintain $M_m = 1.3$. The $G_2(j\omega)$ for $K = 1.3$ and the $M = 1.3$ circle are shown again in Fig. 8-4-3a for reference.

Although many clever methods for compensation by reshaping the Nyquist diagrams have and can be evolved, by and large the process is one of trial and error. This example demonstrates this point.

In the final system $G(j\omega)$ is the product of $G_1(j\omega) \times G_2(j\omega)$. Therefore $G_2(j\omega)$ is reshaped by the phase and amplitude characteristics of $G_1(j\omega)$ to yield the final system. An examination of Fig. 8-4-2b indicates that $G_1(j\omega)$ yields maximum phase shift when y is in the range of 3 to 4. The $G_2(j\omega)$ plot of Fig. 8-4-3a is examined to determine the range of ω that could most profitably be given the maximum phase shift.

In the first portion of this example, it is decided to give $\omega = 1$ on the $G_2(j\omega)$ locus the maximum phase shift. On the $G_1(j\omega)$ locus $y = 3.33$ approximately yields the maximum phase shift. Therefore $G_1(j\omega)$ is unnormalized by the equation

$$y = 3.33 = \frac{\omega}{z_1} = \frac{1}{z_1} \quad (8-4-10)$$

from which z_1 is found as $z_1 = 0.3$. Therefore $G_1(s)$ is

$$G_1(s) = \frac{s + 0.3}{s + 3.0} \quad (8-4-11)$$

The $G(s)$ for the compensated system is

$$G(s) = G_1(s)G_2(s) = \frac{K(s + 0.3)}{s(s + 1)(s + 2)(s + 3)} \quad (8-4-12)$$

The $G(j\omega)$ is plotted and the gain K is set for the $M_m = 1.3$ circle as explained in Art. 8-2. The K is determined as $K = 15$. The final $G(j\omega)$ plot with K set at 15 and the $M = 1.3$ circle are shown in Fig. 8-4-3b. For this system the velocity constant is

$$K_v = \lim_{s \rightarrow 0} sG(s) = 0.75 \quad (8-4-13)$$

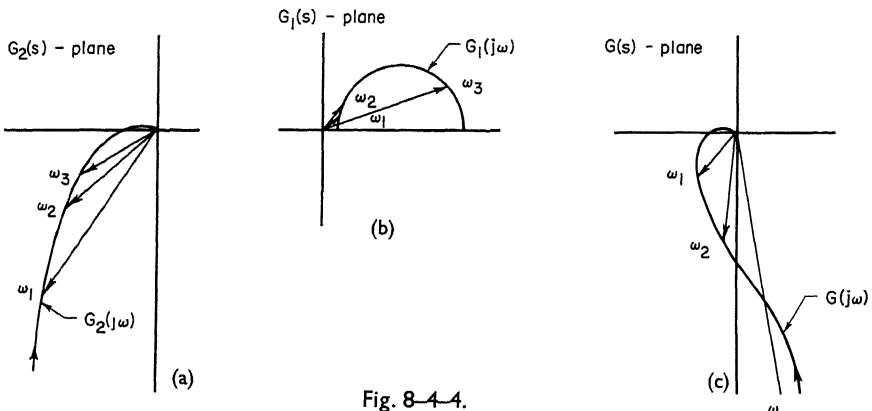


Fig. 8-4-4.

Fig. 8-4-4 emphasizes the graphical interpretation of the compensation process of this example. Fig. 8-4-4a is the locus of the original system with several specific

values of ω indicated. Figure 8-4-4b is the locus of the phase lead network with the same values of ω indicated. Figure 8-4-4c is the compensated system. The phasor indicated by ω_1 for the compensated system is the product of the phasor labeled ω_1 in Fig. 8-4-4a times the phasor labeled ω_1 in Fig. 8-4-4b. Similarly in Fig. 8-4-3c the phasor labeled ω_2 is the product of the ω_2 phasor in Fig. 8-4-4a times the ω_2 phasor in Fig. 8-4-4b, etc.

This graphical interpretation of the compensation process enables the designer to inspect the locus of the uncompensated system and the locus of the compensation network and to visualize the locus of the product without carrying out the actual numerical steps.

Even though K has been increased to 15 from 1.3, the velocity constant has increased only to 0.75 from 0.65. Calculating the velocity constant the compensation network yields the factor $\lim_{s \rightarrow 0} G_1(s) = 0.1$ and K for the entire system must be increased by a factor of more than 10 to increase K_v .

An examination of the final curve of Fig. 8-4-3b indicates that the phase shift that makes $G(j\omega)$ swing to the right of the $-j$ -axis is to no avail. Therefore, as a second try it seems reasonable to have the maximum phase shift of $G_1(j\omega)$ occur at a higher frequency. An inspection of the original $G_2(j\omega)$ plot of Fig. 8-4-3a indicates that $\omega = 3$ might be a reasonable second try. $G_1(p)$ is unnormalized as

$$y = 3.33 = \frac{\omega}{z_1} = \frac{3}{z_1} \quad (8-4-14)$$

from which z_1 is found as $z_1 = 0.9$. Since the entire procedure is an approximation, z_1 is arbitrarily increased to 1 and $G_1(s)$ becomes

$$G_1(s) = \frac{(s + 1)}{(s + 10)} \quad (8-4-15)$$

The $G(s)$ for the compensated system is

$$G(s) = G_1(s)G_2(s) = \frac{K}{s(s + 2)(s + 10)} \quad (8-4-16)$$

The zero in $G_1(s)$ at -1 is said to cancel the pole in $G_2(s)$ at -1 .

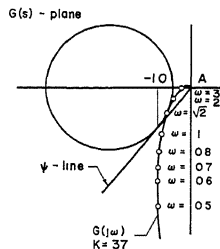


Fig. 8-4-5.

The $G(j\omega)$ is plotted and the gain K is set for the $M_m = 1.3$ circle and K is determined as $K = 37$. The final $G(j\omega)$ plot with K set at 37 and the $M = 1.3$ circle are shown in Fig. 8-4-5. For this system the velocity constant is

$$K_v = \lim_{s \rightarrow 0} sG(s) = 1.85 \quad (8-4-17)$$

The addition of the lead network has increased K , from 0.65 to 1.85, yet the criterion of $M_m = 1.3$ is maintained.

The trial-and-error method can be continued, but examination of the curves of this example indicates that not much can be gained by an additional trial. With a little experience, it is possible to be close to the maximum value of K , in one try.

Other uses of the phase lead network are examined in Art. 8-8.

The network that realizes the $G_1(s)$ of Eq. 8-4-15 can be determined from Eqs. 8-4-3 as

$$z_1 = 1 = \frac{1}{R_1 C_1} \quad (8-4-18)$$

and

$$p_1 = 10 = \frac{R_1 + R_2}{R_1 R_2 C_1} \quad (8-4-19)$$

With two equations in three unknowns, one of the elements can be chosen arbitrarily, and in this example R_1 is set equal to 9×10^5 ohms. With this choice, the other two elements are found as $R_2 = 1 \times 10^5$ ohms and $C_1 = 1.11 \mu\text{f}$.

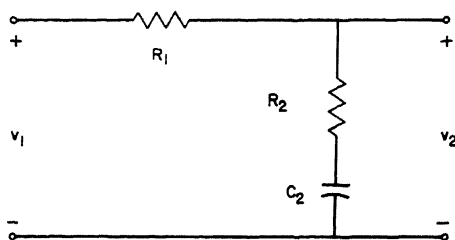


Fig. 8-5-1.

8-5. Phase Lag Network. Another of the common networks used for compensation is the phase lag network shown in Fig. 8-5-1. The transfer function for this network can be written

$$G_1(s) = \frac{V_2}{V_1} = \frac{R_2 C_2 s + 1}{(R_1 + R_2) C_2 s + 1} \quad (8-5-1)$$

or

$$G_1(s) = \frac{V_2}{V_1} = \left[\frac{R_2}{R_1 + R_2} \right] \left[\frac{s + 1/R_2 C_2}{s + 1/C_2(R_1 + R_2)} \right] \quad (8-5-2)$$

With the following definitions

$$z_1 = \frac{1}{R_2 C_2} \quad ; \quad p_1 = \frac{1}{C_2(R_1 + R_2)} \quad (8-5-3)$$

the transfer function can be written

$$G_1(s) = \frac{V_2}{V_1} = \frac{p_1}{z_1} \frac{(s + z_1)}{(s + p_1)} \quad (8-5-4)$$

From Eqs. 8-5-3, the ratio of z_1 to p_1 is

$$\frac{z_1}{p_1} = \frac{R_1 + R_2}{R_2} \quad (8-5-5)$$

Therefore the zero is always farther from the origin than the pole, and the magnitude of this ratio can be changed by varying R_1 and R_2 . If this ratio is set too high, the high frequencies may be attenuated by too large an amount. Therefore this ratio is often given the value 10, and with this arbitrary choice $G_1(s)$ becomes

$$G_1(s) = \frac{V_2}{V_1} = \frac{s + z}{10(s + 0.1z)} \quad (8-5-6)$$

The change of variables

$$p = \frac{s}{z_1} = \frac{\sigma}{z_1} + j\frac{\omega}{z_1} = x + jy \quad (8-5-7)$$

is used to normalize $G_1(s)$ as

$$G_1(p) = \frac{p + 1}{10(p + 0.1)} \quad (8-5-8)$$

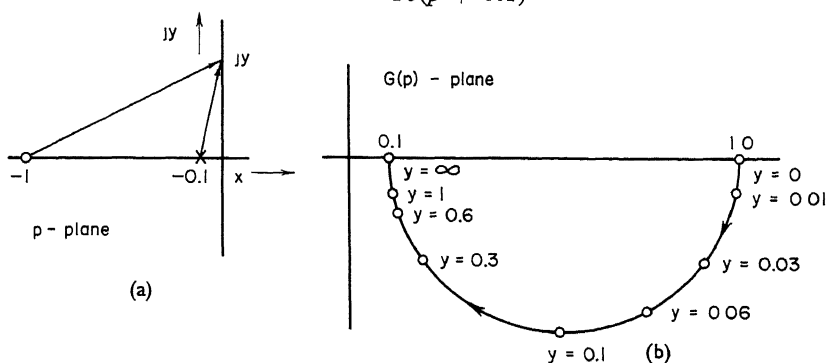


Fig. 8-5-2.

A plot of $G_1(jy)$ can be obtained from this equation. Again as s moves up the $j\omega$ -axis, p moves up the jy -axis as shown in Fig. 8-5-2a. Since the pole is inside the zero, the phasor in the denominator swings faster than the phasor in the numerator and the result is the lagging phase characteristics of Fig. 8-5-2b.

Example 8-5-1. The system

$$G_2(s) = \frac{K}{s(s+1)(s+2)} \quad ; \quad H(s) = 1 \quad (8-5-9)$$

is used again. It is desired to compensate this system with a lag network in order to increase the velocity constant and yet maintain $M_m = 1.3$.

An inspection of the $G_1(jy)$ curve of Fig. 8-5-2b shows that, for higher values of y , $G_1(jy)$ is essentially equal to 0.1. If y is unnormalized so that this portion of the curve corresponds to the ω values where $G_2(j\omega)$ is close to the -1 point, then $G_1(j\omega)G_2(j\omega)$ is essentially 0.1 of $G_2(j\omega)$ alone. Therefore, the gain and the velocity constant can be increased by 10.

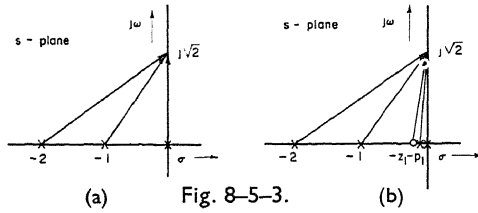


Fig. 8-5-3.

Another way of looking at this situation is as follows: The poles of $G_2(s)$ are shown in Fig. 8-5-3a. When $j\omega = j\sqrt{2}$, the phase angle of $G_2(j\omega)$ is -180° . In Fig. 8-5-3b the pole and zero of the lag network are added and it is seen that if they are close to the origin, the effective phase angle of $G_1(j\omega)$ is essentially zero when $j\omega = j\sqrt{2}$. Therefore, the only effect on $G(j\omega) = G_1(j\omega)G_2(j\omega)$ in this range of ω is to multiply $G_2(j\omega)$ by 0.1. Therefore, for the same relative stability, the system gain can be increased by 10, and this in turn increases K_v by 10.

Because the element sizes for the network that realizes $G_1(s)$ become larger and larger as the pole and zero of $G_1(s)$ are moved closer and closer to the origin, there are practical considerations that limit the procedure, and the entire gain of 10 can not be realized.

To continue the example, $G_1(s)$ is chosen as

$$G_1(s) = \frac{s + 0.1}{10(s + 0.01)} \quad (8-5-10)$$

and $G(s)$ becomes

$$G(s) = G_1(s)G_2(s) = \frac{K(s + 0.1)}{10(s)(s + 1)(s + 2)(s + 0.01)} = \frac{K'(s + 0.1)}{s(s + 1)(s + 2)(s + 0.01)} \quad (8-5-11)$$

The constant K' is defined to absorb the 0.1 factor.

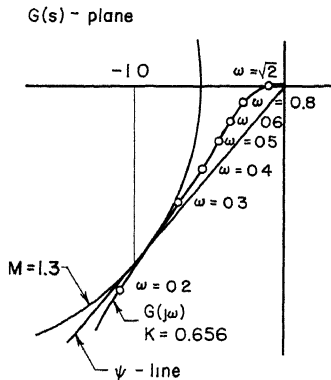


Fig. 8-5-4.

The $G(j\omega)$ is plotted and the gain K' is set for the $M_m = 1.3$ circle and K' is determined as $K' = 0.656$. The final $G(j\omega)$ locus with K' set at 0.656 and the $M = 1.3$ circle are shown in Fig. 8-5-4. For this system, the velocity constant is

$$K_v = \lim_{s \rightarrow 0} sG(s) = 3.28 \quad (8-5-12)$$

The addition of the lag network has increased K_v from 0.65 to 3.28, yet the criterion of $M_m = 1.3$ is maintained. The lag network is discussed again in Art. 8-13.

The network that realizes the $G_1(s)$ of Eq. 8-5-10 can be determined from Eqs. 8-5-3 as

$$z_1 = 0.1 = \frac{1}{R_2 C_2} \quad (8-5-13)$$

and

$$p_1 = 0.01 = \frac{1}{(R_1 + R_2)C_2} \quad (8-5-14)$$

Again one element value is arbitrary and R_1 is chosen as $R_1 = 9 \times 10^5$ ohms. The other two elements are found as $R_2 = 1 \times 10^5$ ohms and $C_2 = 100 \mu\text{f}$.

Even in this example the element values are becoming large as can be seen by $C_2 = 100 \mu\text{f}$. It might be wise in the final design to increase both R_1 and R_2 by some factor so that C_2 is decreased.

8-6. Lag-Lead Network. Another of the common networks used for compensation is the lag-lead network shown in Fig. 8-6-1. The transfer function for this network can be written

$$G_1(s) = \frac{V_2}{V_1} = \frac{(s + 1/R_1 C_1)(s + 1/R_2 C_2)}{s^2 + (1/R_1 C_1 + 1/R_2 C_2 + 1/R_2 C_1)s + 1/R_1 R_2 C_1 C_2} \quad (8-6-1)$$

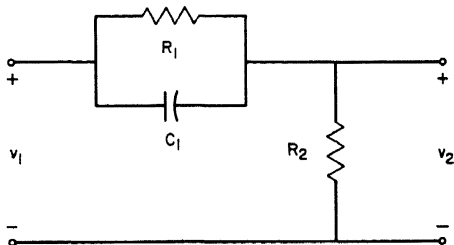


Fig. 8-6-1.

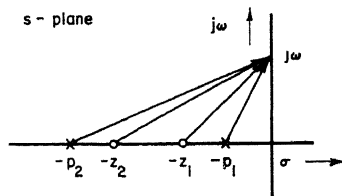


Fig. 8-6-2.

By inspection it can be seen that the zeros for this transfer function are at $s = -1/R_1 C_1$ and $s = -1/R_2 C_2$. It can be seen also that the product of the poles is equal to the product of the zeros but that the magnitude of the sum of the poles is larger than the magnitude of the sum of the zeros. From this it can be reasoned that the two zeros are between the two poles, as shown in Fig. 8-6-2. At low frequencies, the pole-zero combination closer to the origin makes the network behave as if it were a lag network. At higher frequencies, the pole-zero combination farther out causes the network to act as if it were a lead network. A more exact analysis follows:

The lag-lead network has zero phase shift at zero and infinite frequencies, but it also has zero phase shift at an intermediate frequency. This intermediate frequency occurs when the angle of the denominator equals the angle of the numerator in $G_1(j\omega)$. This frequency is determined from the equation

$$G_1(j\omega) = \frac{(-\omega^2 + 1/R_1C_1R_2C_2) + j\omega(1/R_1C_1 + 1/R_2C_2)}{(-\omega^2 + 1/R_1C_1R_2C_2) + j\omega(1/R_1C_1 + 1/R_2C_2 + 1/R_2C_1)} \quad (8-6-2)$$

By inspection it can be seen that the angle of the numerator and the angle of the denominator are both $+90^\circ$ when the real part of each is equal to zero or when

$$\omega_1 = \frac{1}{\sqrt{R_1R_2C_1C_2}} \quad (8-6-3)$$

When $\omega = \omega_1$, then $G_1(j\omega)$ becomes

$$G_1(j\omega) = \frac{1/R_1C_1 + 1/R_2C_2}{1/R_1C_1 + 1/R_2C_2 + 1/R_2C_1} = \frac{R_1C_1 + R_2C_2}{R_1C_1 + R_2C_2 + R_1C_2} \quad (8-6-4)$$

This is the minimum value of $G_1(j\omega)$. Again in order not to have some frequencies attenuated by too large a value, this minimum value of $G_1(j\omega)$ is arbitrarily set as 0.1

$$G_1(j\omega_1) = \frac{R_1C_1 + R_2C_2}{R_1C_1 + R_2C_2 + R_1C_2} = 0.1 \quad (8-6-5)$$

The pole-zero pattern of Eq. 8-6-1 can be written as

$$G_1(s) = \frac{(s + z_1)(s + z_2)}{(s + p_1)(s + p_2)} \quad (8-6-6)$$

where

$$z_1 = \frac{1}{R_1C_1} \quad ; \quad z_2 = \frac{1}{R_2C_2} \quad (8-6-7)$$

and the p 's can be determined from the denominator of Eq. 8-6-1.

Because the number of variables is large, another arbitrary choice can be made. For this development the ratio of z_2 to z_1 is chosen as 5.

Eq. 8-6-6 becomes

$$G_1(s) = \frac{(s + z_1)(s + 5z_1)}{(s + p_1)(s + p_2)} \quad (8-6-8)$$

Determining z_2 in terms of z_1 effectively determines the ω at which the magnitude of $G_1(j\omega)$ is minimum in terms of z_1 . This is shown by starting with Eq. 8-6-3 as

$$\omega_1 = \frac{1}{\sqrt{R_1R_2C_1C_2}} = \sqrt{z_1z_2} = \sqrt{5z_1} \quad (8-6-9)$$

Next, the pole locations of Eq. 8-6-8 are determined. Eq. 8-6-5 in terms of p 's and z 's becomes

$$G_1(j\omega_1) = \frac{z_1 + z_2}{p_1 + p_2} = \frac{z_1 + 5z_1}{p_1 + p_2} = 0.1 \quad (8-6-10)$$

It must also be remembered that the product of the poles is equal to the product of the zeros

$$p_1 p_2 = z_1 z_2 = 5z_1^2 \quad (8-6-11)$$

When p_2 is eliminated between Eq. 8-6-10 and Eq. 8-6-11, the resulting equation in p_1 can be written as

$$p_1^2 - 60z_1 p_1 + 5z_1^2 = 0 \quad (8-6-12)$$

The roots of this equation are approximately

$$p_1 = 0.0833z_1 \text{ and } 59.2z_1 \quad (8-6-13)$$

For convenience, the second root is approximated as $60z_1$. Since p_1 and p_2 in Eq. 8-6-8 are symmetrical, one of the values can be chosen for p_1 and the other for p_2 . Finally, Eq. 8-6-8 can be written as

$$G_1(s) = \frac{(s + z_1)(s + 5z_1)}{(s + 0.0833z_1)(s + 60z_1)} \quad (8-6-14)$$

Eq. 8-6-14 can be normalized by the substitution

$$p = \frac{s}{z_1} = \frac{\sigma}{z_1} + j\frac{\omega}{z_1} = x + jy \quad (8-6-15)$$

and the equation becomes

$$G_1(p) = \frac{(p + 1)(p + 5)}{(p + 0.0833)(p + 60)} \quad (8-6-16)$$

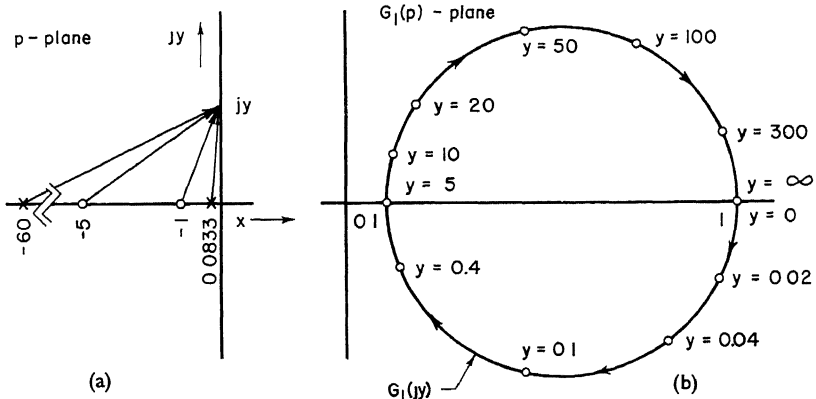


Fig. 8-6-3.

A plot of $G_1(jy)$ can be obtained from this equation. Again as s moves up the $j\omega$ -axis, p moves up the jy -axis as shown in Fig. 8-6-3a. For obvious reasons this figure is not drawn to scale. The resulting locus is shown in Fig. 8-6-3b. An inspection of this figure indicates that at low frequencies the network has a lagging phase characteristic, while at high frequencies it has a leading phase characteristic. Hence the name lag-lead.

The discussion for the lag-lead network is reviewed briefly. The transfer function for the lag-lead network has two poles and two zeros.

Only three of the four quantities are independent as the product of the poles must equal the product of the zeros. When the magnitude of $G_1(j\omega)$ is set at 0.1, only two of the four quantities are independent. When the ω at which $G_1(j\omega) = 0.1$ is set, only one of the four quantities is independent. The remaining quantity can be varied as needed in design.

Obviously, several arbitrary choices have been made in this development. Under certain situations, it may be wise to make other choices. However, the procedure remains the same.

Example 8-6-1. The system

$$G_2(s) = \frac{K}{s(s+1)(s+2)} ; \quad H(s) = 1 \quad (8-6-17)$$

is used again. It is desired to compensate this system with a lag-lead network in order to increase the velocity constant and yet maintain $M_m = 1.3$.

The $G_2(j\omega)$ locus for the uncompensated system with $K = 1.3$ and the $M_m = 1.3$ circle is shown in Fig. 8-6-4a.

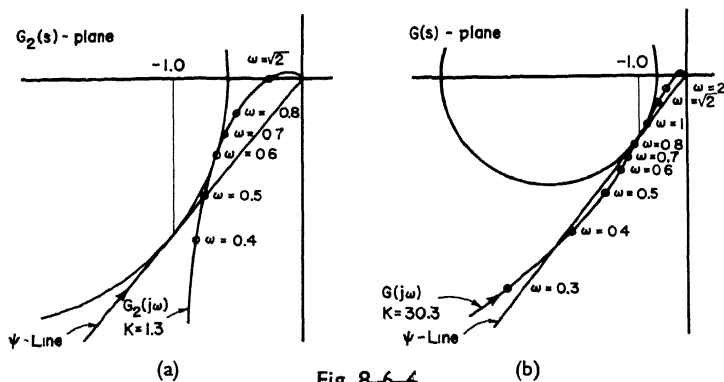


Fig. 8-6-4.

The $G_1(jy)$ curve is unnormalized by the following cut-and-try procedure. From Fig. 8-6-3b, it can be seen that when $y = \sqrt{5}$, $G_1(jy)$ has no phase shift. The corresponding $G_1(j\omega)$ also offers no phase shift at the corresponding ω , and $G(j\omega) = G_1(j\omega)G_2(j\omega)$ moves in and out on a radial line through some point on the $G_2(j\omega)$ locus as K is changed. At frequencies below this, the $G(j\omega)$ will be shifted in the lagging direction, and at frequencies above this in the leading direction.

Fig. 8-6-4a is now examined with this point in mind. It can certainly be seen that the ω of $G_2(j\omega)$ corresponding to zero phase shift should not be in the range of $\omega = 1$. The $G(j\omega)$ curve at frequencies below $\omega = 1$ will be shifted in the lagging direction and these curves will intersect with the $M = 1.3$ curve for low values of K .

At the other extreme, it can be seen that the ω of $G_2(j\omega)$ corresponding to zero phase shift should not be in the range $\omega = 0.1$. Therefore from inspection of $G_2(j\omega)$, the zero phase shift ω should be in the range of 0.4 to 0.5. In this example $\omega = 0.45$ is used and $G_1(p)$ is unnormalized as

$$y = \sqrt{5} = \frac{\omega}{z_1} = \frac{0.45}{z_1} \quad (8-6-18)$$

from which z_1 is found approximately as $z_1 = 0.2$. Therefore $G_1(s)$ becomes

$$G_1(s) = \frac{(s + 0.2)(s + 1)}{(s + 0.0166)(s + 12)} \quad (8-6-19)$$

The $G(s)$ for the compensated system is

$$G(s) = G_1(s)G_2(s) = \frac{K(s + 0.2)}{s(s + 0.0166)(s + 2)(s + 12)} \quad (8-6-20)$$

The $G(j\omega)$ is plotted and the gain K is set for the $M_m = 1.3$ circle and K is determined as $K = 30.3$. The final $G(j\omega)$ locus with K set at 30.3 and the $M = 1.3$ circle are shown in Fig. 8-6-4b. For this system the velocity constant is

$$K_v = \lim_{s \rightarrow 0} sG(s) = 15.15 \quad (8-6-21)$$

Other trials might increase the K_v somewhat, but an examination of the curves of this example indicate that this K_v is not far from the maximum.

TABLE 8-7-1

Compensation network	Velocity constant	ω at which M_m occurs
None	0.65	0.6
Lead	1.85	1.5
Lag	3.28	0.25
Lag-lead	15.15	0.95

8-7. Comparison of Results. The results of the previous four examples are summarized in Table 8-7-1. The column for ω_m (the ω at which M_m occurs) is added in order to stress the point that other factors may be important in a design problem in addition to M_m .

The lead network is essentially a high-pass filter and ω_m occurs at a higher frequency than that of the uncompensated system.

The lag network is essentially a low-pass filter and ω_m occurs at a lower frequency than that for the uncompensated system.

The lag-lead network is essentially a band elimination filter; but ω_m is on the high side of the minimum attenuation point; therefore ω_m is increased slightly compared with the uncompensated system.

Frequently, because of noise and other unwanted signals, it is desired to maintain the system bandwidth as small as possible without undue attenuation of the signal itself. Therefore ω_m is also of interest to the designer.

8-8. Compensating the Type 2 System. The usual Type 0 and Type 1 systems which have Nyquist diagrams originating for $\omega = 0$ on the $+$ real axis or at $-j\infty$, respectively, can be made stable by reducing the gain sufficiently. If compensation is used, it is for the purpose of

improving some aspect of the design, as by an increase in velocity constant. However, a Type 2 system can be unstable for all values of system gain. Therefore compensation has the more fundamental requirement of producing a stable system.

A specific example of a Type 2 system that is unstable for all values of K is given by

$$G_2(s) = \frac{K(s + z_1)}{s^2(s + p_1)} \quad (8-8-1)$$

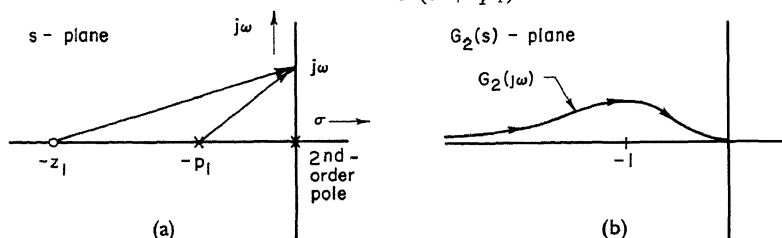


Fig. 8-8-1.

where the pole at $-p_1$ is inside the zero at $-z_1$, as shown in Fig. 8-8-1a. The $G_2(j\omega)$ locus is shown in Fig. 8-8-1b.

Adding a lag network to this system is of no avail. The resulting pole-zero pattern is shown in Fig. 8-8-2. The pole and zero of the lag network are labeled $-p_n$ and $-z_n$ respectively. The original system already suffers from the fact that $-p_1$ is closer to the origin than $-z_1$, and the lag network with its pole closer to the origin than its zero only exaggerates this lagging characteristic.

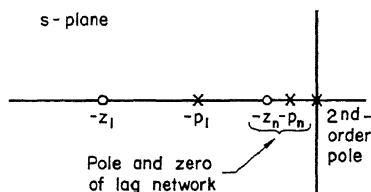


Fig. 8-8-2.

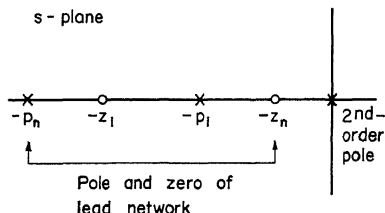


Fig. 8-8-3.

The only possible compensation that can stabilize this system is one that has leading phase characteristics so that the entire $G(j\omega)$ locus is moved to the other side of the -1 point. A lead network with the zero inside the pole is such a network. The resulting pole-zero pattern of a lead network added to the original system is shown in Fig. 8-8-3. If $-z_n$, the zero of the lead network, is added inside $-p_1$ and $-p_n$, the pole outside $-z_1$, the resulting $G(j\omega)$ certainly depicts a stable system. This can be seen from the fact that $-z_n$ and $-p_1$ when considered as a pair produce a net leading phase as do $-z_1$ and $-p_n$ when considered as a pair.

The zero $-z_n$ does not have to be inside the pole. If, for example, $-z_n$ is outside $-p_1$ but close to $-p_1$, then $-z_n$ and $-p_1$ essentially cancel and the system is effectively left with $-z_1$ inside $-p_n$, thus yielding

a stable system. The only requirement that must be met is that the $G(j\omega)$ must be shifted to the other side of the -1 point from that of $G_2(j\omega)$.

If a Type 2 system is sufficiently stable, a lag network may be used for reasons other than absolute stability. The $G_2(j\omega)$ for such a system is indicated in Fig. 8-8-4. The effect of adding a lag network will shift the $G(j\omega)$ toward the -1 point but the resulting system may still have the desired relative stability. Since a Type 2 system has an infinite velocity constant, the lag network is not added to improve K_v but it may be used to increase the acceleration constant.

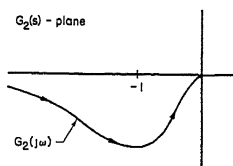


Fig. 8-8-4.

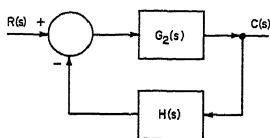
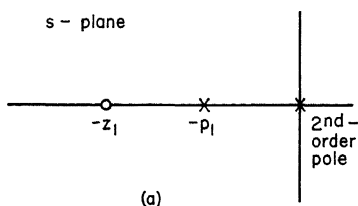
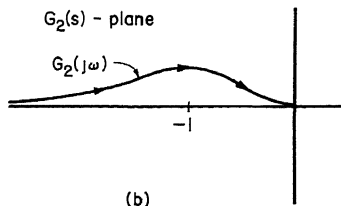


Fig. 8-9-1.

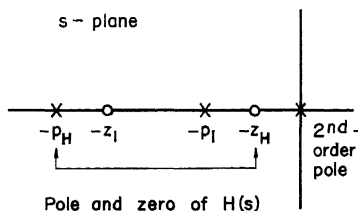
8-9. Feedback Compensation. A system can be compensated by adding a network (or some other compensating device) in the feedback path. Such a system is shown in Fig. 8-9-1. As far as absolute stability is concerned, the locus plotted for the Nyquist diagram is $H(j\omega)G_2(j\omega)$. Therefore a $G_2(j\omega)$ locus that indicates an absolutely unstable system with unity feedback can be reshaped by adding a proper $H(s)$ in the feedback path. The pole-zero pattern for such a system is shown in Fig. 8-9-2a and a sketch of $G_2(j\omega)$ is shown in Fig. 8-9-2b indicating an absolutely unstable system. A pole and a zero of an $H(s)$ are added to the



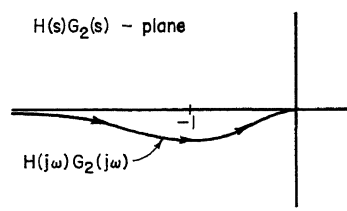
(a)



(b)



(c)



(d)

Fig. 8-9-2.

system as depicted in Fig. 8-9-2c and the resulting $H(j\omega)G_2(j\omega)$ locus of Fig. 8-9-2d demonstrates that the system has been stabilized.

The frequency response of the closed-loop system cannot be obtained by adding the M and N circles directly to Fig. 8-9-2d. However, as explained in Chapter 7, the transfer function for the closed-loop system can be written as

$$\frac{C(s)}{R(s)} = \left[\frac{H(s)G_2(s)}{1 + H(s)G_2(s)} \right] \frac{1}{H(s)} \quad (8-9-1)$$

The M and N circles can be added to this figure to obtain the magnitude and phase of

$$\frac{H(j\omega)G_2(j\omega)}{1 + H(j\omega)G_2(j\omega)} \quad (8-9-2)$$

but these results must be modified by the factor $1/H(j\omega)$ to yield information about $C(j\omega)/R(j\omega)$. The designer can observe the shapes of these two curves, and by a trial-and-error procedure he can adjust the system for any desired value of M_m .

In other situations an auxiliary feedback path may be used to reshape the locus of a particular element or group of elements that form a portion

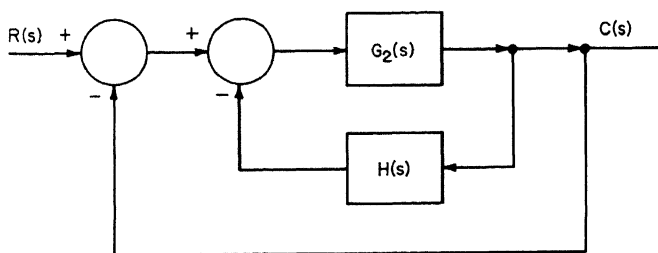


Fig. 8-9-3.

of a more elaborate system as shown in Fig. 8-9-3. In this system it is assumed that $G_2(j\omega)$ indicates an undesirable system and that the $G_2(j\omega)$ curve needs to be reshaped.

A specific example is used to explain the situation. It is assumed that $G_2(s)$ is given by

$$G_2(s) = \frac{K}{s(s + p)} \quad (8-9-3)$$

as can be seen, $G_2(s)$ by itself could lead to a Type 1 system with an infinite position constant.

As a first step in exploring this situation, $H(s)$ is assumed to be unity. The inner loop becomes the $G(s)$ for the over-all system as given by

$$G(s) = \frac{G_2(s)}{1 + G_2(s)} = \frac{K}{s^2 + ps + K} \quad (8-9-4)$$

When the over-all loop is closed around this $G(s)$, the final system becomes a Type 0 system with a finite position constant. If the over-all system is to maintain an infinite position constant, the $H(s)$ cannot be unity.

With $H(s)$ added, the transfer function of the inner loop becomes

$$G(s) = \frac{G_2(s)}{1 + H(s)G_2(s)} \quad (8-9-5)$$

If for some range of values of $s = j\omega$, the product $H(s) \times G_2(s)$ is small compared with unity, Eq. 8-9-5 can be approximated as

$$G(s) \approx G_2(s) \quad (8-9-6)$$

If, for some other range of values of $s = j\omega$, the product $H(s)G_2(s)$ is large compared with unity, Eq. 8-9-5 can be approximated as

$$G(s) \approx \frac{G_2(s)}{H(s)G_2(s)} = \frac{1}{H(s)} \quad (8-9-7)$$

In order to maintain the over-all system as a Type 1 system, the pole at the origin in $G_2(s)$ must also appear in $G(s)$. An examination of Eq. 8-9-6 shows that, if the $\lim_{s \rightarrow 0} H(s)G_2(s) = 0$, the pole at the origin in $G_2(s)$ does appear in $G(s)$. The necessary characteristics of $H(s)$ can be seen from

$$\lim_{s \rightarrow 0} H(s)G_2(s) = \lim_{s \rightarrow 0} \frac{H(s)K}{s(s + p_1)} = 0 \quad (8-9-8)$$

For Eq. 8-9-8 to hold, $H(s)$ must have at least a second-order zero at the origin. One possible $H(s)$ is given by

$$H(s) = \frac{s^2(s + z_1)}{(s + p_1)(s + p_2)(s + p_3)} \quad (8-9-9)$$

A network that yields this $H(s)$ can be synthesized by methods explained in Chapter 12.

When the product $H(s)G_2(s)$ is large, Eq. 8-9-7 shows that $G(s)$ is approximately equal to $1/H(s)$ and the characteristics of $G(s)$ are dominated by $H(s)$. The designer then has the ability to reshape this portion of the $G(j\omega)$ curve by his choice of $H(s)$.

8-10. Discussion of Compensation. Up to this point, three types of networks and the manner in which they are used in series compensation have been discussed. Parallel compensation has also been considered briefly.

Since series compensation has been featured more than parallel compensation, this discussion centers around the examples based on series compensation. However, in general terms, the discussion applies to all forms of compensation.

Three compensation networks are featured in the discussion of series compensation. In addition to these three, there are an infinite number of other networks that can be used. Although there are occasional exceptions, the network parameters generally must be adjusted by a trial-and-error procedure. After this the gain is set by the graphical process for a desired value of M_m . No matter how many networks are analyzed, the designer has the concern that some other network might yield more satisfactory results.

The difficulties of designing a system by reshaping the Nyquist diagram are particularly apparent when other factors such as rise time, overshoot, and bandwidth are also to be controlled. For these reasons, other design methods have been developed, of which the root-locus method is explained in Chapter 11 and a direct synthesis procedure in Chapter 13.

8-11. Nyquist Diagram from Experimental Measurements.

Frequently the performance of an existing closed-loop system may be unsatisfactory, or in some situations the designer may hesitate to close the loop on a system being designed without first checking the system experimentally. In either case if the open-loop system is stable, the loop is opened and a signal of fixed amplitude but variable frequency is applied as the actuating signal. By some convenient method, both the amplitude and the phase of the output signal versus frequency are measured. If the system has elements in the feedback path, $B(j\omega)$ is also measured.

For the present discussion, it is assumed that $H(s) = 1$. The $G(j\omega)$ curve resulting from these experimental data can be plotted. From this plot, the designer has available information about absolute stability and relative stability. If the indicated stability is not satisfactory, a compensation device can be added by using the techniques explained in the previous articles.

8-12. Comparison of Transient and Frequency Response Terminology for the Lead Network. In Chapter 6, the terms “error-rate damping” and “integral compensation” are used, whereas in this chapter such terms as “lead-and-lag” compensation are used. It is instructive to investigate the fact that compensation with a lead network approximates error-rate compensation whereas compensation with a lag network approximates integral compensation. The lead network is examined first.

In the proportional system, the signal applied to the remainder of the system is proportional to the actuating signal as

$$m(t) = Ke(t) \quad (8-12-1)$$

For convenience, this time signal is called $m(t)$ and the transform of the signal $M(s)$. The transform of Eq. 8-12-1 is

$$M(s) = KE(s) \quad (8-12-2)$$

In the proportional plus error-rate system, the signal applied to the remainder of the system is the sum of two terms one of which is proportional to the actuating signal and the other proportional to the derivative of the actuating signal as

$$m(t) = Ke(t) + K_1 \frac{de(t)}{dt} \quad (8-12-3)$$

The transform of this equation is

$$M(s) = KE(s) + K_1 sE(s) \quad (8-12-4)$$

The lead network of Fig. 8-4-1 has the transfer function

$$\frac{V_2(s)}{V_1(s)} = \frac{R_2 + R_1 R_2 C_1 s}{R_1 + R_2 + R_1 R_2 C_1 s} \quad (8-12-5)$$

The next step is mathematically unprecise, but is included to introduce a concept that is explored more rigorously in terms of a range of frequencies. If

$$R_1 + R_2 \gg R_1 R_2 C_1 s \quad (8-12-6)$$

then the denominator of Eq. 8-12-5 can be approximated by $R_1 + R_2$, and

$$\frac{V_2(s)}{V_1(s)} \approx \frac{R_2 + R_1 R_2 C_1 s}{R_1 + R_2} \quad (8-12-7)$$

The equation is solved for $V_2(s)$ and written as

$$V_2(s) \approx \frac{R_2}{R_1 + R_2} V_1(s) + \frac{R_1 R_2 C_1 s}{R_1 + R_2} V_1(s) \quad (8-12-8)$$

The time equation corresponding to this is

$$v_2(t) \approx \frac{R_2}{R_1 + R_2} v_1(t) + \frac{R_1 R_2 C_1}{R_1 + R_2} \frac{dv_1(t)}{dt} \quad (8-12-9)$$

This time equation indicates why the lead network is sometimes called a differentiating network. Actually the term "proportional-plus-derivative" would be a more descriptive title.

When Eq. 8-12-9 is compared with Eq. 8-12-3 or when Eq. 8-12-8 is compared with Eq. 8-12-4, the similarity between the two sets of equations can be seen. In the completed system, the term $R_2/(R_1 + R_2)$ is only one of the factors of K , and similarly $R_1 R_2 C_1/(R_1 + R_2)$ is only one of the factors of K_1 . However, it can be seen that by the proper choice of R_1 , R_2 , and C_1 , K , and K_1 can be made to have any desired ratio.

As stated, the inequality

$$R_1 + R_2 \gg R_1 R_2 C_1 s \quad (8-12-10)$$

needs to be explored because of the s factor. More significance can be given this equation in terms of sinusoidal steady-state conditions. The s in

Eq. 8-12-5 can be replaced by $s = j\omega$ to yield

$$\frac{V_2(j\omega)}{V_1(j\omega)} = \frac{R_2 + j\omega R_1 R_2 C_1}{R_1 + R_2 + j\omega R_1 R_2 C_1} \quad (8-12-11)$$

If ω is sufficiently small, the denominator can be approximated by $R_1 + R_2$. As an arbitrary choice, the ratio 10:1 between the real and the j components is used to determine $\omega_{\max}$:

$$R_1 + R_2 = 10R_1 R_2 C_1 \omega_{\max} \quad (8-12-12)$$

From this it can be seen that the lead network tends to act as a differentiating device for ω in the range

$$0 < \omega < \omega_{\max} = \frac{R_1 + R_2}{10R_1 R_2 C_1} \quad (8-12-13)$$

Frequency plots are used to summarize this discussion. The transfer function for the idealized error-rate damping is given by

$$\frac{M(s)}{E(s)} = K + K_1 s \quad (8-12-14)$$

which in terms of frequency becomes

$$\frac{M(j\omega)}{E(j\omega)} = K + j\omega K_1 \quad (8-12-15)$$

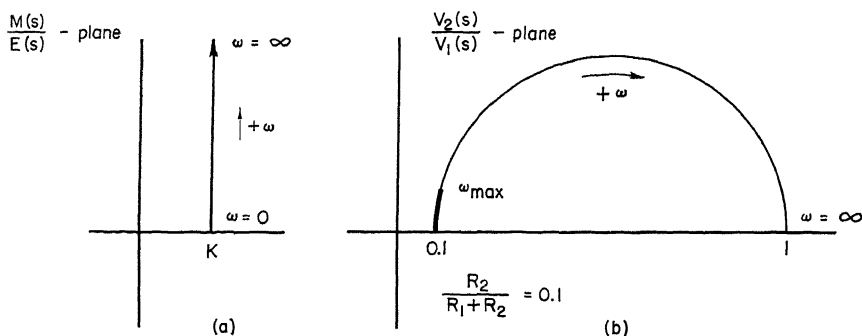


Fig. 8-12-1.

The frequency locus for this equation is shown in Fig. 8-12-1a where the locus starts at $+K$ on the real axis with $\omega = 0$ and moves parallel to the jV -axis as ω increases. The frequency response of the actual lead network is shown in Fig. 8-12-2b under the assumption that $R_2/(R_1 + R_2) = 0.1$. This plot is a semicircle as discussed in Art. 8-4. The $\omega_{\max}$ from Eq. 8-12-12 is shown also. A comparison of the two curves indicates that in the range $0 < \omega < \omega_{\max}$ that the semicircle of Fig. 8-12-1b does approximate the straight line of Fig. 8-12-1a.

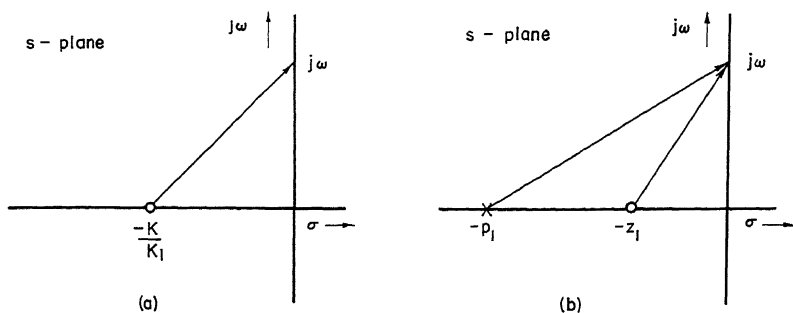


Fig. 8-12-2.

Finally, before leaving the subject, the pole-zero configurations for the two systems are compared. The idealized error-rate damping system has only a zero, as shown in Fig. 8-12-2a. In order that there be a zero with an *RC* network, there must also be an accompanying pole. About the best that can be done is to place the pole farther out from the origin than the zero, as shown in Fig. 8-12-2b.

An examination of Fig. 8-12-2a indicates that as ω moves up the $j\omega$ -axis, the angle on the phasor from the zero begins at zero and increases to $+90^\circ$. An examination of Fig. 8-12-2b shows that for small values of ω , the angle on the phasor from $-p_1$, is small and contributes little to the total angle of the lead network. Therefore, for small values of ω , the idealized system is approximated by the lead network. However, for large ω , the angle on the phasor from $-p_1$, cannot be ignored, and the behavior of the actual lead network and that of the idealized system differ greatly. Therefore, for ω in the range from 0 to $\omega_{\max}$, the lead network approximates the idealized system.

8-13. Comparison of Transient and Frequency Response Terminology for the Lag Network. In the proportional-plus-integral compensation system, the signal applied to the remainder of the system is of the form

$$m(t) = Ke(t) + K_2 \int e(t) dt \quad (8-12-1)$$

The equation is transformed as

$$M(s) = KE(s) + \frac{K_2}{s}E(s) \quad (8-13-2)$$

The transfer function is

$$\frac{M(s)}{E(s)} = K + \frac{K_2}{s} \quad (8-13-3)$$

The lag network of Fig. 8-5-1 has the transfer function that can be put in the form

$$\frac{V_2(s)}{V_1(s)} = \frac{R_2 + 1/sC_2}{R_1 + R_2 + 1/sC_2} \quad (8-13-4)$$

Again the next step is mathematically unprecise but is included to introduce the concept that follows. If

$$R_1 + R_2 \gg \frac{1}{sC_2} \quad (8-13-5)$$

then Eq. 8-13-4 can be approximated by

$$\frac{V_2(s)}{V_1(s)} \approx \frac{R_2}{R_1 + R_2} + \frac{R_2}{sC_2(R_1 + R_2)} \quad (8-13-6)$$

The equation is solved for $V_2(s)$:

$$V_2(s) \approx \frac{R_2}{R_1 + R_2} V_1(s) + \frac{R_2}{C_2(R_1 + R_2)} \frac{V_1(s)}{s} \quad (8-13-7)$$

The corresponding time equation is

$$v_2(t) \approx \frac{R_2}{R_1 + R_2} v_1(t) + \frac{R_2}{C_2(R_1 + R_2)} \int_0^t v_1(t) dt \quad (8-13-8)$$

This time equation indicates why the lag network is sometimes called an integrating network.

When Eq. 8-13-8 is compared with Eq. 8-13-1, or when Eq. 8-13-7 is compared with Eq. 8-13-2, the similarity between the two sets of equations is clear. By the proper choice of R_1 , R_2 , and C_2 , the designer can cause K and K_2 to have any desired ratio.

As stated, the inequality

$$R_1 + R_2 \gg \frac{1}{sC_2} \quad (8-13-9)$$

needs to be explored because of the s factor. This is done by writing Eq. 8-13-4 with s replaced by $j\omega$

$$\frac{V_2(j\omega)}{V_1(j\omega)} = \frac{R_2 + 1/j\omega C_2}{R_1 + R_2 + 1/j\omega C_2} \quad (8-13-10)$$

If ω is sufficiently large, the denominator can be approximated by $R_1 + R_2$. As an arbitrary choice, the ratio 10:1 between the real and the j components is used to determine $\omega_{\min}$ as

$$R_1 + R_2 = \frac{10}{C_2 \omega_{\min}} \quad (8-13-11)$$

From this it can be seen that the lag network tends to act as an integrating device for ω in the range

$$\omega_{\min} = \frac{10}{(R_1 + R_2)C_2} < \omega < \infty \quad (8-13-12)$$

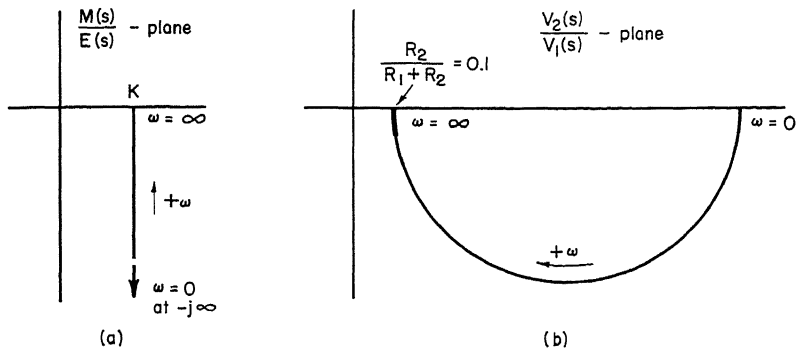


Fig. 8-13-1.

Frequency plots are used to summarize this discussion. The transfer function of Eq. 8-13-3 in terms of frequency becomes

$$\frac{M(j\omega)}{E(j\omega)} = K + \frac{K_2}{j\omega} \quad (8-13-13)$$

The frequency locus for this equation is shown in Fig. 8-13-1a where the locus moves up a line parallel to the $-jV$ -axis. When $\omega = 0$, the locus starts at $-j\infty$, and when $\omega = \infty$, the locus terminates at $+K$ on the real axis. The frequency response of the actual lag network is shown in Fig. 8-13-1b under the assumption that $R_2/(R_1 + R_2) = 0.1$. This plot is a semicircle as discussed in Art. 8-5. The $\omega_{\min}$ of Eq. 8-13-12 is also shown. A comparison of the two curves shows that in the range $\omega_{\min} < \omega < \infty$ that the semicircle of Fig. 8-13-1b does approximate the straight line of Fig. 8-13-1a.

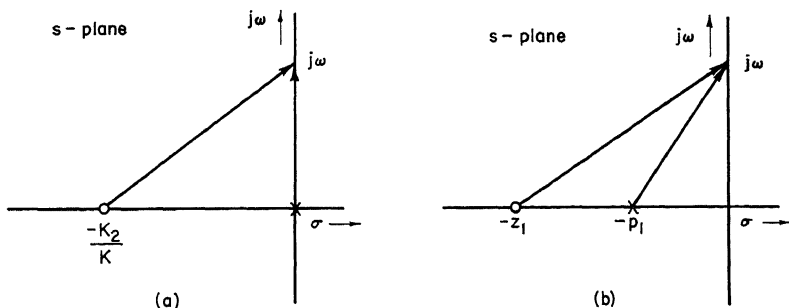


Fig. 8-13-2.

Finally, the pole-zero configurations for the two systems are compared. The transfer function for the idealized integral compensation can be written

$$\frac{M(s)}{E(s)} = K + \frac{K_2}{s} = \frac{K(s + K_2/K)}{s} \quad (8-13-14)$$

The pole-zero configuration for this is shown in Fig. 8-13-2a. The pole-zero configuration for the lag network is shown in Fig. 8-13-2b. The lag network cannot create the pole at the origin of the idealized system, and therefore about the best that can be done is to bring the pole at $-p_1$ as close to the origin as is practicable.

An examination of Fig. 8-13-2a shows that for small values of ω the angle on the phasor from the pole at the origin is always 90° . However from Fig. 8-13-2b, it can be seen that for small values of ω the phasor from the pole at $-p_1$ has a small angle. Therefore, for small values of ω , the behavior of the actual network and that of the idealized system differ greatly. When ω moves up the $j\omega$ -axis, however, the phasor from the pole at $-p_1$ begins to approach 90° and the two systems begin to behave in a similar manner. Therefore, for ω larger than $\omega_{\min}$, the lag network approximates the idealized system.

PROBLEMS

8-1. For the system

$$G(s) = \frac{K}{s(s+1)(s+2)} ; \quad H(s) = 1$$

set the gain K for M_m to have the following values:

- (a) $M_m = 1.1$
- (b) $M_m = 1.4$
- (c) $M_m = 2.0$

8-2. For the system

$$G(s) = \frac{K}{s(s+3)} ; \quad H(s) = 1$$

set the gain K for M_m to have the following values:

- (a) $M_m = 1.3$
- (b) $M_m = 2.0$
- (c) $M_m = 5.0$

8-3. For the system

$$G(s) = \frac{K}{s(s+3)(s+6)} ; \quad H(s) = 1$$

set the gain K for M_m to have the values:

- (a) $M_m = 1.1$
- (b) $M_m = 1.4$
- (c) $M_m = 2.0$

After determining K by graphical means, scale the $G(s)$ by the change in variables by $s = 3p$ and compare the values of K found in this problem with those of Prob. 8-1.

8-4. In Art. 8-2 it is suggested that, in setting the gain K , the value $K = 1$ be initially given to plot $G(j\omega)$. Develop the equation for determining K (analogous to Eq. 8-2-5) if K is initially given any arbitrary value.

8-5. In plotting Fig. 8-4-2b, the ratio of z_1 to p_1 is arbitrarily set at 1:10 and the resulting plot of $G_1(j\omega)$ is shown in Fig. 8-4-2b. The maximum value of leading phase for this figure can be measured as approximately $\Phi_{\max} = 53^\circ$. Change the ratio of z_1 to p_1 in the following steps: 1:40, 1:20, 1:8, 1:6, 1:4, and 1:2. For each ratio determine φ_m by using graphical procedures. Compare the results in a table.

8-6. For the system of Example 7-7-4 repeated as

$$G(s) = \frac{K(s+4)}{s(s-1)} \quad ; \quad H(s) = 1$$

develop the procedure so that K can be set for

(a) $M_m = 5.0$

(b) $M_m = 2.0$

(c) $M_m = 1.2$

8-7. Suppose it is desired to set the gain K of a system for a certain M_m . When $G(j\omega)$ is plotted and the ψ line is added, the two curves coincide for some distance. This means that any number of circles can be drawn tangent to both curves in the process of setting the gain. Explain the significance of this situation in terms of design.

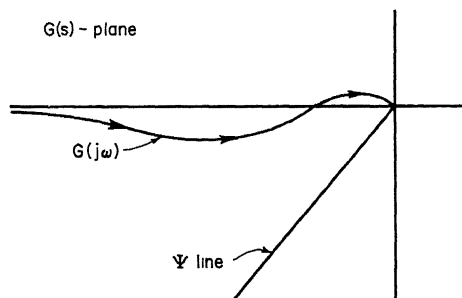


Fig. 8-P-8.

8-8. A Type 2 system has a $G(j\omega)$ locus as shown in Fig. 8-P-8, and it is desired to set the gain for a certain value of M_m . The resulting ψ line is drawn but it is discovered to be impossible to add a circle tangent to both curves. Explain the significance of this situation.

8-9. A system is known to be unstable and yet it is possible to draw a circle tangent to both the $G(j\omega)$ and a ψ line determined from a desired value of M_m . Sketch such a $G(j\omega)$ and ψ line. What is the significance of this situation?

8-10. Given a system

$$G(s) = \frac{K}{s(s+1)^2} \quad ; \quad H(s) = 1$$

It is desired to set K for $M_m = 1.4$. After determining K , also determine K_v .

8-11. For the system of Prob. 8-10, it is desired to increase K_v as much as possible by adding a lead network and yet to maintain $M_m = 1.4$. Determine

$G_1(s)$, the transfer function of the network that will do this, assuming the ratio of z_1 to p_1 to be 1 to 10. From $G_1(s)$ determine R_2 and C_1 when $R_1 = 9 \times 10^5$ ohms.

8-12. For the system of Prob. 8-10, it is desired to increase K_v as much as possible by adding a lag network and yet to maintain $M_m = 1.4$. $G_1(s)$ is chosen as

$$G_1(s) = \frac{(s + 0.1)}{10(s + 0.01)}$$

Determine K_v for the resulting system.

8-13. Repeat Prob. 8-12 for

$$G_1(s) = \frac{(s + 0.01)}{10(s + 0.001)}$$

Determine K_v and compare with the K_v of Prob. 8-12. Also determine the necessary R_2 and C_2 where $R_1 = 9 \times 10^5$ ohms and compare C_1 with that required in Prob. 8-12.

8-14. The system of Prob. 8-10 is to be compensated with a lag-lead network where $G_1(s)$ is given by Eq. 8-6-14. Maintain $M_m = 1.4$ and increase K_v as much as possible. In the resulting network, determine R_2 , C_1 , and C_2 when $R_1 = 9 \times 10^5$ ohms.

8-15. In Ex. 8-6-1, determine R_2 , C_1 , and C_2 necessary for the circuit to yield the $G_1(s)$ of Eq. 8-6-19 when $R_1 = 9 \times 10^5$ ohms.

8-16. Repeat Probs. 8-10 through 8-14 as assigned for

$$G(s) = \frac{K(s + 1.5)}{s(s + 1)(s + 2)} \quad ; \quad H(s) = 1$$

8-17. Repeat Prob. 8-16 for

$$G(s) = \frac{K(s + 4)}{s(s + 1)(s + 2)} \quad ; \quad H(s) = 1$$

8-18. Given a system

$$G(s) = \frac{K(s + 0.9)}{s^2(s + 1)} \quad ; \quad H(s) = 1$$

It is desired to set K for $M_m = 1.4$. Investigate whether this is possible, and if it is, determine K .

8-19. If the value of $M_m = 1.4$ in Prob. 8-18 cannot be obtained by varying K , the system is to be compensated by using a lead network which has the ratio of z_1 to p_1 equal to 1 to 10. If the system can be so compensated, adjust the network parameters to obtain a maximum acceleration constant.

8-20. Explore the possibility of compensating the system of Prob. 8-18 with a lag network.

8-21. Explore the possibility of compensating the system of Prob. 8-18 with a lag-lead network.

8-22. Develop the graphical techniques necessary to set the gain K for a desired M_m circle by using the inverse complex plane plot.

8-23. In Art. 8-11, the use of Nyquist diagrams from experimental measurements is discussed. The loop is opened and a signal of fixed amplitude and variable

frequency is applied as the actuating system. Explore this concept in the following situation. The closed-loop system is a stable Type 2 system with unity feedback. The test signal applied to $G(s)$ is $e(t) = \sin \omega t$. Explore the nature of the $e(t)$ response, giving particular attention to the effect of the double pole at the origin in $G(s)$. Next apply a test signal $e(t) = \cos \omega t$. Explain the reason for the different type of response due to the double pole at the origin in $G(s)$.

8-24. Repeat Prob. 8-23 for a Type 3 system.

FREQUENCY RESPONSE CHARACTERISTICS

9-1. General Frequency Response. In this age of stereophonic and hi-fi sound, most engineers are familiar with the frequency-response characteristic of an amplifier. This characteristic shows the steady-state sinusoidal amplification ratio or gain for the amplifier over a suitable frequency range. Generally the gain is plotted according to a rather loose interpretation of decibels (db) defined as

$$\text{number of db} = 20 \log_{10} |A(j\omega)| \quad (9-1-1)$$

where $A(j\omega)$ is the expression for the steady-state sinusoidal amplification. In order to show the gain over a wide frequency range, a logarithmic frequency scale is used. A typical gain curve is shown in Fig. 9-1-1a.

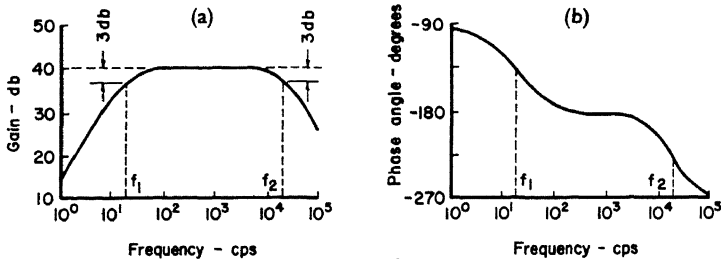


Fig. 9-1-1.

Since $A(j\omega)$ has both magnitude and phase, the gain curve of Fig. 9-1-1a should be supplemented by a phase curve as in Fig. 9-1-1b. Curves of the types shown in Fig. 9-1-1 are the basis of a whole area of study in control systems. The technique is applied to various control-system transfer functions in order to determine such things as absolute and relative stability, peak amplification, resonances, etc. The ease with which frequency-response curves of this type can be sketched from basic transfer functions, and then modified as compensation is added, is a most attractive feature.

The frequency-response graphs contain no more information about transfer functions than do the polar or Nyquist plots. However the ease of sketching transfer functions in the form of Fig. 9-1-1 makes such sketches more desirable than the Nyquist diagrams in many cases.

This chapter considers the details of sketching the frequency response

plots which are commonly referred to in the control system field as Bode diagrams. Following this study of diagram construction some of the various aspects of using Bode plots in analysis and design are considered.

9-2. Bode Diagram Gain Plots. The various elements or parts of a control system have frequency-response characteristics as do electronic amplifiers. In fact, most control components behave as various forms of low-pass filters. In order to construct the Bode gain curves for a part of a system, the decibel relation of Eq. 9-1-1 is applied to the desired steady-state transfer function which is designated as $G(j\omega)$:

$$\text{db gain} = 20 \log_{10} |G(j\omega)| \quad (9-2-1)$$

If $G(j\omega)$ can be written in the form of a ratio of polynomials such as

$$G(j\omega) = \frac{z_1 z_2 K}{p_1 p_2} \frac{(j\omega/z_1 + 1)(j\omega/z_2 + 1)}{(j\omega/p_1 + 1)(j\omega/p_2 + 1)} \quad (9-2-2)$$

then

$$\begin{aligned} \text{db gain} = & 20 \log_{10} \left| \frac{K z_1 z_2}{p_1 p_2} \right| + 20 \log_{10} \left| j\frac{\omega}{z_1} + 1 \right| + 20 \log_{10} \left| j\frac{\omega}{z_2} + 1 \right| \\ & - 20 \log_{10} |j\omega| - 20 \log_{10} \left| j\frac{\omega}{p_1} + 1 \right| - 20 \log_{10} \left| j\frac{\omega}{p_2} + 1 \right| \end{aligned} \quad (9-2-3)$$

This process changes products and quotients of factors into sums and differences, thereby accomplishing one simplification. All numerator factors have plus signs while all denominator factors have negative signs. It should also be noted that the gain constant $|K z_1 z_2 / p_1 p_2|$ simply adds a certain constant number of decibels that is independent of frequency. If interest is centered in the gain variation with frequency, the gain constant can be forgotten temporarily and only factors of the form $\pm 20 \log_{10} |j\omega|$ or $\pm 20 \log_{10} |j\omega/z + 1|$ need be considered.

As a first step in construction, the factor

$$\text{db gain} = 20 \log_{10} |j\omega| = 20 \log_{10} \omega \quad (9-2-4)$$

is examined. For a $\log_{10}$ frequency scale and a linear decibel scale, Eq. 9-2-4 can be written as

$$y = 20x \quad (9-2-5)$$

where y is the number of decibels and x is the logarithm of frequency (angular velocity). Thus, on a plot of linear x and y scales, Eq. 9-2-5 is a straight line. The actual calculation of $\log_{10} \omega$ can be omitted by simply using semilog graph paper, as shown in Fig. 9-2-1. Two things should be noticed about the line for the term $j\omega$ in Fig. 9-2-1. At $\omega = 1$, the decibel gain is zero, or the gain ratio is one. Second, the slope of the line is 20 db per decade, or for every factor of 10 increase in frequency, the

gain changes by 20 decibels. The low frequencies are spread out more than the high frequencies. In many cases the slope of the $|j\omega|$ curve in Fig. 9-2-1 is given in terms of an octave of frequency, which is a change by a ratio of 2 or $\frac{1}{2}$. To determine the slope on this basis, a frequency ω_1 is chosen and the change in decibel gain is calculated in going to $2\omega_1$. The particular value of ω_1 makes no difference.

$$\text{Change in db} = 20 \log_{10} 2\omega_1 - 20 \log_{10} \omega_1 = 20 \log_{10} 2 \approx 6 \text{ db} \quad (9-2-6)$$

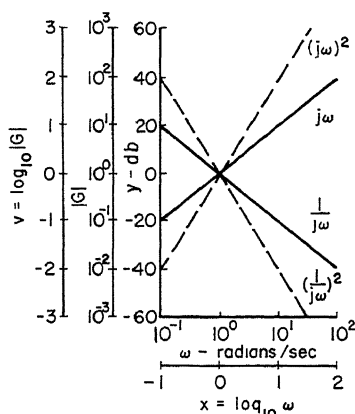


Fig. 9-2-1.

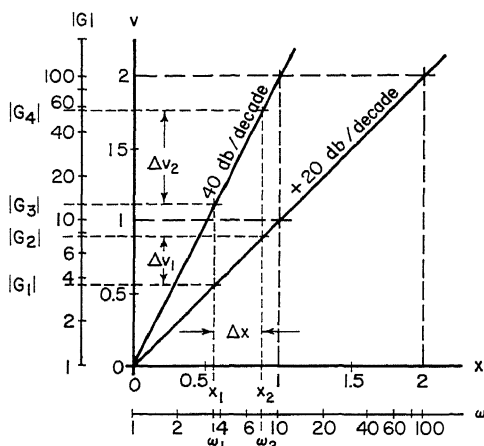


Fig. 9-2-2.

Thus a transfer function term of the form $j\omega$ can be represented by a line with a slope of 6db/octave or 20db/decade.

Fig. 9-2-1 includes an additional gain ratio scale. This scale points out the possibility of using a log axis vertically as well as horizontally. Thus log-log graph paper can be used if desired. This eliminates the need to calculate gain ratios from decibel values when reading from the curves. Examples that follow show the use of both log-log and semilog graph paper.

A short digression into some analytics of log coordinates is in order. This discussion is undertaken with the graph in Fig. 9-2-2. It is frequently necessary to divide a segment of the x -axis into parts (x and v are termed exponential units), or to find the relationship in ω that represents an increment Δx . From Fig. 9-2-2

$$\Delta x = x_2 - x_1 = \log \omega_2 - \log \omega_1 = \log \frac{\omega_2}{\omega_1} \quad (9-2-7)$$

Thus a given frequency ratio is always represented by the same increment or linear distance on the horizontal axis. The same must be true of gain ratios on the vertical axis. A common problem might be to divide the decade of frequency from $\omega = 10$ to $\omega = 100$ into quarters. The increment Δx is thus 0.25 and the corresponding frequency ratio is

$$0.25 = \log \frac{\omega_2}{\omega_1} \quad (9-2-8)$$

$$\frac{\omega_2}{\omega_1} = \sqrt[4]{10} = 1.78 \quad (9-2-9)$$

Thus with $\omega_1 = 10$, the point $x = 1.25$ is found to be $(1.78)(10) = 17.8$ radians/sec. Similarly $x = 1.5$ is found either by

$$\frac{\omega_2}{17.8} = 1.78 \quad ; \quad \omega_2 = 31.6 \quad (9-2-10)$$

or from

$$0.5 = \log \frac{\omega_2}{10} \quad ; \quad \omega_2 = 10\sqrt{10} \quad (9-2-11)$$

A second useful property of log coordinates relates to slopes of common curves as shown in Fig. 9-2-2. For the +20 db/decade sloping line it is apparent that the slope on the x and v scales is +1. Similarly a +40 db/decade sloping line is shown. Its corresponding slope for x and v coordinates is +2. Easy relations now give corresponding values of Δv_1 and Δv_2 . For the +1 sloping line

$$\Delta v_1 = \Delta x \quad (9-2-12)$$

$$\log |G_2| - \log |G_1| = \log \omega_2 - \log \omega_1 \quad (9-2-13)$$

$$\frac{|G_2|}{|G_1|} = \frac{\omega_2}{\omega_1} \quad (9-2-14)$$

Similarly for the +2 sloping line

$$\log |G_4| - \log |G_3| = 2(\log \omega_2 - \log \omega_1) = \log \left(\frac{\omega_2}{\omega_1} \right)^2 \quad (9-2-15)$$

$$\frac{|G_4|}{|G_3|} = \left(\frac{\omega_2}{\omega_1} \right)^2 \quad (9-2-16)$$

This scheme can be extended such that

$$\frac{|G_j|}{|G_i|} = \left(\frac{\omega_j}{\omega_i} \right)^m \quad (9-2-17)$$

where m is the slope (+ or -) of the line segment in (x, v) coordinates. It is thus advisable to become familiar with slopes in terms of both decibels and exponents of the logarithmic base in use. The particular base 10 is not required; however, it is very convenient. Further use of the above relations is pointed out later.

A term more commonly encountered than $j\omega$ is $1/j\omega$, which leads to

$$\text{number of db} = -20 \log_{10} \omega \quad (9-2-18)$$

The only difference from Eq. 9-2-4 is a minus sign or a negative slope to the curve, as shown in Fig. 9-2-1. The line representing this type of frequency-response term goes through zero db at $\omega = 1$ and has a slope of -20db/decade or -6db/octave.

Some control-system steady-state transfer functions have terms of the form $(j\omega)^2$ or $1/(j\omega)^2$. These lead to expressions such as

$$\text{number of db} = \pm 20 \log_{10} \omega^2 = \pm 40 \log_{10} \omega \quad (9-2-19)$$

The Bode curves are still straight lines which go through zero decibels at $\omega = 1$, but they have slopes twice as great as have the lines for $j\omega$ and $1/j\omega$. The slopes are 40 db/decade or 12 db/octave. In fact any frequency response terms of the form $(j\omega)^n$ or $1/(j\omega)^n$ can be represented on a Bode plot by straight lines. They go through zero decibels at $\omega = 1$ with slopes of $\pm 20n$ db/decade, or $\pm 6n$ db/octave, or simply $\pm n$ exponential units.

While considering the simplest frequency-varying response term, it is advisable to introduce the effect of frequency-invariant gain K which was dropped earlier in this article. A term of the form $K/(j\omega)^n$, when expressed in decibels, gives

$$\text{number of db} = 20 \log_{10} K - 20n \log_{10} \omega \quad (9-2-20)$$

Thus a constant number of decibels $20 \log_{10} K$ is added to the frequency-varying term. At $\omega = 1$ the line now goes through the point $20 \log_{10} K$. In other words, the line is raised a fixed amount. Its slope is not changed. The sole effect of adding the constant K is to slide the line up or down from the $K = 1$ position by the amount $20 \log_{10} K$. In changing a Bode plot from a gain K_1 to a gain K_2 the frequency-varying curves are slid up or down an amount $20 \log_{10}(K_2/K_1)$.

An alternate way of accounting for a gain K is to slide the gain or decibel scales up or down without changing the line itself. Thus a relative gain or decibel scale assuming $K = 1$ can be used to make the first sketch and the actual or absolute decibel scale can be inserted to account for the actual gain ratio K .

Example 9-2-1. A Bode plot for the transfer function

$$G(s) = 40/s^2 \quad (9-2-21)$$

is to be plotted first on semilog paper and then on log-log paper. For the semilog plot $20 \log_{10} 40 = 32$ db is needed. The curve slopes downward at 40 db/decade and goes through 32 db at $\omega = 1$. Fig. 9-2-3a shows the required curve plus a

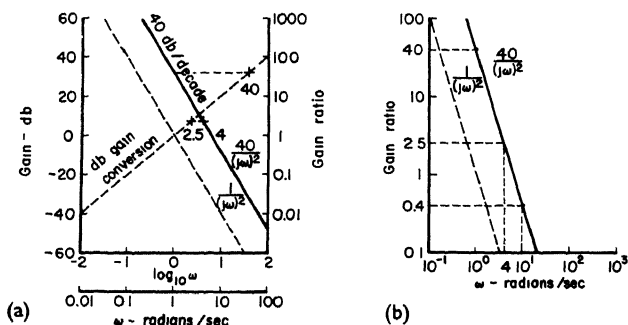


Fig. 9-2-3.

curve for $K = 1$ to illustrate the shift due to K . Fig. 9-2-3b shows the log-log plot with only a gain-relative-to-1 vertical scale. In drawing this curve it should be noted that, as ω increases 10 times, the $|G(j\omega)|$ falls by 10^2 times, and at $\omega = 1$, $|G(j1)| = 40$. Two points suitable for completing this curve are shown in Fig. 9-2-3b. The vertical scale on the log-log plot cannot be adjusted as can a decibel scale on a semilog plot. The number of log cycles more or less fixes the size of the vertical scale on the log-log plot, while on the semilog plot the linear vertical divisions can be scaled as desired. The log-log plot allows actual gain to be read directly at any frequency, as at $\omega = 4$, $|G(j4)| = 2.5$; while the semilog plot gives $|G(j4)| = 7.96$ db, which must be converted to numerical gain by an antilog process, $|G(j4)| = 10^{7.96/20} = 2.5$. A convenient aid in this process is the graph in Fig. 9-2-4, which can be used for converting decibels to actual numerical gain. This graph can be superimposed on a Bode plot as shown in Fig. 9-2-3a, thereby allowing the numerical gain to be read from the horizontal scale. In this example at $\omega = 4$, the gain in decibels is found from the $40/(j\omega)^2$ curve, projected horizontally to the decibel conversion curve, and the numerical gain read on the horizontal scale as 2.5.

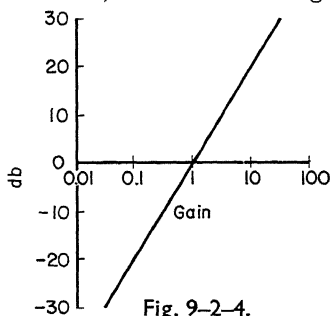


Fig. 9-2-4.

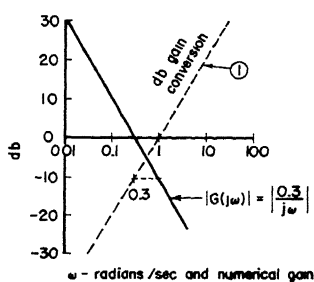


Fig. 9-2-5.

Example 9-2-2. As another example a Bode plot for

$$G(s) = 0.3/s \quad (9-2-22)$$

is to be constructed on semilog paper and a decibel-to-numerical-gain conversion curve superimposed on it. The conversion curve can be drawn first by choosing an arbitrary zero decibel line and scale and numbering the horizontal scale. Curve ① of Fig. 9-2-5 results. To draw the Bode curve, the following steps are followed. At 0.3 on the horizontal scale the corresponding number of decibels, -10.5 , is read on the conversion curve and projected horizontally to the $\omega = 1$ line. The Bode curve goes through this point with a slope of -20 db/decade. Of course the conversion curve can be omitted and $20 \log_{10} (0.3)$ calculated directly. The student should get used to the significance of negative numbers of decibels which only mean numerical gains less than 1.

The common factor to be considered next on the Bode plot is of the form $|j\omega/z + 1|$ from Eq. 9-2-3. In terms of decibels this term results in

$$\text{number of db} = 20 \log_{10} \left[\left(\frac{\omega}{z} \right)^2 + 1 \right]^{1/2} \quad (9-2-23)$$

In order to simplify sketching the Bode plot of this function various frequency regions are considered separately. For frequencies much less than $\omega = z$, Eq. 9-2-23 can be approximated as

$$\text{number of db} \approx 20 \log_{10} 1 = 0 \quad ; \quad \omega \ll z \quad (9-2-24)$$

Actually, if $\omega \leq z/3$, the approximation is valid. Eq. 9-2-24 is simply the line corresponding to zero decibels. At frequencies much greater than $\omega = z$, Eq. 9-2-24 can be approximated as

$$\text{number of db} \approx 20 \log_{10} \left(\frac{\omega}{z} \right) \quad ; \quad \omega \gg z \quad (9-2-25)$$

This type of variation is similar to the curve for $G(j\omega) = j\omega$. However Eq. 9-2-25 gives zero decibels at $\omega = z$. Thus Eq. 9-2-25 plots as a straight line through the point $\omega = z$ with 20 db/decade slope. It must be remembered that this curve is valid only for values of ω much greater than z . By combining the lines for Eqs. 9-2-24 and 9-2-25, an asymptotic approximation to the Bode plot for Eq. 9-2-23 is obtained. Fig. 9-2-6 shows the results. The break in the lines occurs at $\omega = z$, which is called a break frequency. For frequencies near $\omega = z$, the straight lines are in error and corrections need to be computed. At $\omega = z$, Eq. 9-2-23 becomes

$$\text{number of db} = 20 \log_{10}[1 + 1] \approx 3 \text{ db} \quad (9-2-26)$$

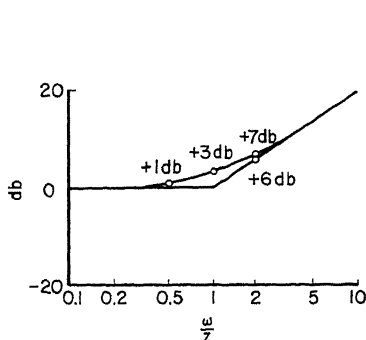


Fig. 9-2-6.

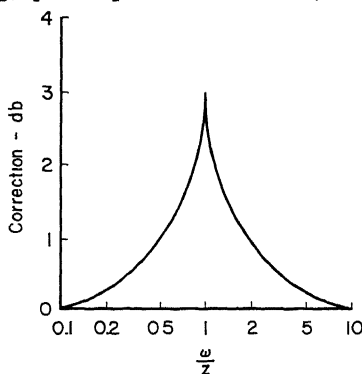


Fig. 9-2-7.

indicating a correction of 3 db at the break frequency. Similar corrections of 1 db at frequencies an octave above and below the break frequency are very useful, also. More corrections than these three are rarely needed; however, Fig. 9-2-7 shows corrections for any desired frequency relative to $\omega = z$. These are corrections to be added to the number of decibels given by the straight-line asymptotes.

When a term of the form $|1/(j\omega/p + 1)|$ occurs, the reasoning applied above can easily be extended to give the Bode plot. In decibels the relation becomes

$$\text{number of db} = -20 \log_{10} \left[\left(\frac{\omega}{p} \right)^2 + 1 \right]^{1/2} \quad (9-2-27)$$

The asymptote for values of ω much less than p is still the zero decibel line while the asymptote for values of ω much greater than p is given by

$$\text{number of db} \approx -20 \log_{10}(\omega/p) \quad ; \quad \omega \gg p \quad (9-2-28)$$

Thus the sloping asymptote has a negative slope in this case, or the break is down at the break frequency $\omega = p$. The correction curve of Fig. 9-2-7 can be used for plotting a more exact Bode plot if the values are read as negative values and ω/z is replaced by ω/p . Fig. 9-2-8 illustrates the Bode plot for this case, which is merely the negative of the previous case.

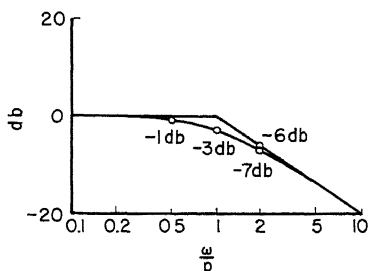


Fig. 9-2-8.

When $G(j\omega)$ has terms such as $(j\omega/z + 1)^n$ or $(j\omega/p + 1)^{-n}$ a slight extension of the above asymptotic construction techniques must be accomplished. The decibel equations become

$$\text{number of db} = \pm 20n \log_{10} \left[\left(\frac{\omega}{p} \right)^2 + 1 \right]^{1/2} \quad (9-2-29)$$

While the basic straight lines and break frequencies are not changed, the slope of the high-frequency asymptotes is changed from ± 20 db/decade to $\pm 20n$ db/decade or $\pm 6n$ db/octave. The correction curve also must be multiplied by n .

The appearance of a gain constant K with the simple zero or pole factors considered above can be accounted for by the same method used for the pole or zero at the origin. The whole Bode plot is raised or lowered by an amount $20 \log_{10} K$.

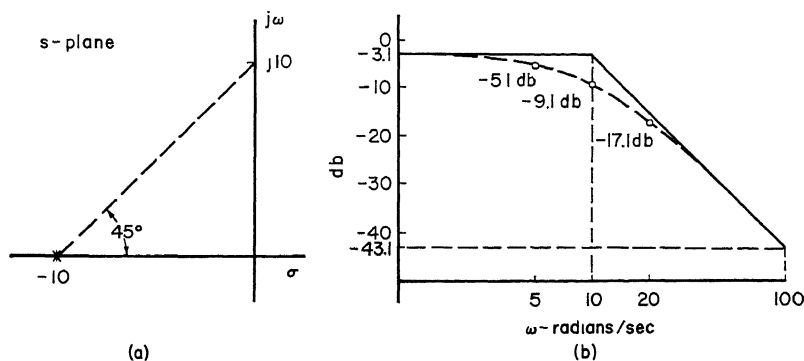


Fig. 9-2-9.

Example 9-2-3. The Bode plot for

$$G(s) = \frac{70}{(s + 10)^2} = \frac{0.7}{\left(\frac{s}{10} + 1\right)^2} \quad (9-2-30)$$

is to be constructed. To relate the Bode plot to the s -plane poles and zeros, Fig. 9-2-9a shows the double pole at $s = -10$ and a geometrical relation that gives the break frequency $\omega = 10$ radians/second. The break downward is at a rate of 20×2 db/decade at this frequency. The presence of the gain 0.7 means that the low-frequency asymptote is the line $20 \log_{10}(0.7) = -3.1$ db instead of zero decibels. Fig. 9-2-9b shows the required plot.

For more complicated transfer functions involving terms of the types treated so far, the Bode plots involve only sums and differences of the graphs for the terms already discussed. The following examples illustrate the general construction procedures.

Example 9-2-4. The Bode plot for

$$G(s) = \frac{100(s + 6)}{s + 60} = 10 \frac{s/6 + 1}{s/60 + 1} \quad (9-2-31)$$

is to be constructed. In terms of the steady-state transfer function and decibels

$$20 \log_{10} |G(j\omega)| = 20 \log_{10} 10 + 20 \log_{10} \left| j\frac{\omega}{6} + 1 \right| - 20 \log_{10} \left| j\frac{\omega}{60} + 1 \right| \quad (9-2-32)$$

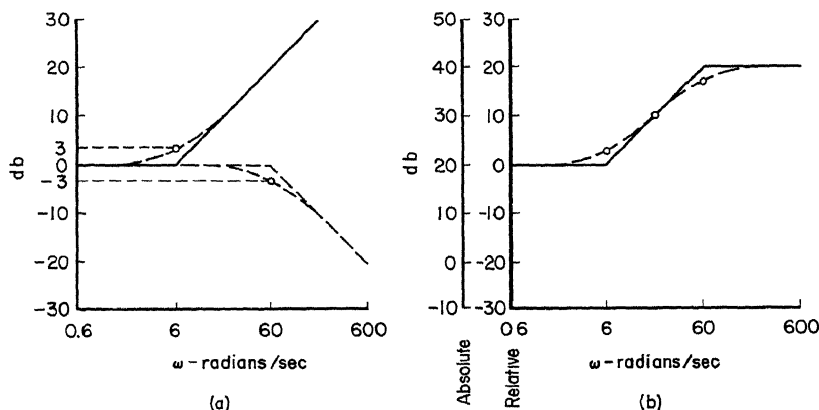


Fig. 9-2-10.

The second and third terms in Eq. 9-2-32 are sketched separately in Fig. 9-2-10a. The sum of these two terms is shown in Fig. 9-2-10b. The gain constant term $20 \log_{10} 10$ shifts the curve of Fig. 9-2-10b up by 20 db. This shift is shown by adding a new absolute decibel scale shifted down as shown in Fig. 9-2-10b. The original scale can be termed a relative decibel scale. Above $\omega = 60$ the positive and negative slopes counteract each other and the curve levels out. The corrections for each original part can be read from Fig. 9-2-7 and a more accurate plot can be made.

Example 9-2-5. As another example, the asymptotic Bode plot for

$$G(s) = \frac{45000(s + 3)}{s(s + 15)(s + 30)^2} \quad (9-2-33)$$

$$G(j\omega) = \frac{10\left(j\frac{\omega}{3} + 1\right)}{j\omega\left(j\frac{\omega}{15} + 1\right)\left(j\frac{\omega}{30} + 1\right)^2} \quad (9-2-34)$$

is to be determined. The various terms give break frequencies at $\omega = 3, 15, 30$ radians/sec. By neglecting the gain constant 10 temporarily the various components can be sketched as shown in Fig. 9-2-11a. The addition of these components

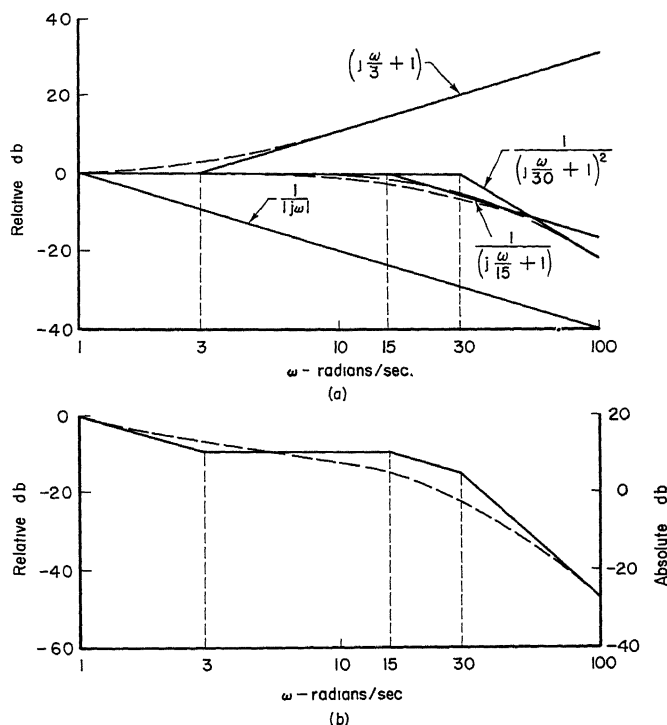


Fig. 9-2-11.

proceeds as follows. For frequencies below $\omega = 3$ the $1/j\omega$ term predominates, with the other terms adding zero decibels to the line representing $1/j\omega$. At $\omega = 3$ the term $j\omega/3 + 1$ becomes significant, and its 20 db/decade positive slope counteracts the -20 db/decade slope of the $1/j\omega$ term, causing the total curve to level out at 0 db/decade. There is no further change until $\omega = 15$, at which frequency the term $1/(j\omega/15 + 1)$ becomes significant, causing the total curve to break down at 20 db/decade. This slope continues until $\omega = 30$ when the factor $1/(j\omega/30 + 1)^2$ adds an additional slope downward of 40 db/decade. Therefore, the final slope above $\omega = 30$ is -60 db/decade. Fig. 9-2-11b shows the result of summing the component curves of Fig. 9-2-11a. The inclusion of $K = 10$ simply shifts the whole

curve upward 20 db. Instead of redrawing the plot, a new absolute decibel scale is added shifted down 20 db. Again, corrections can be included for a more accurate Bode plot by using the correction curve of Fig. 9-2-7 for the terms that break at $\omega = 3, 15$, and 30, and then adding all corrections from the smooth curves in Fig. 9-2-11a.

One additional type of transfer-function factor is of importance in Bode plot construction. This is the quadratic factor

$$s^2 + 2\zeta\omega_n s + \omega_n^2 \quad \text{or} \quad 1/(s^2 + 2\zeta\omega_n s + \omega_n^2).$$

Since the Bode plot for one of these factors is simply the negative of the other, only the term $s^2 + 2\zeta\omega_n s + \omega_n^2$ is considered. The steady-state transfer function for this term can be arranged as

$$G(j\omega) = K \left[\left(1 - \frac{\omega^2}{\omega_n^2} \right) + j2\zeta \left(\frac{\omega}{\omega_n} \right) \right] \quad (9-2-35)$$

If K is neglected temporarily, this becomes

$$\text{number of db} = 20 \log_{10} \left[\left(1 - \frac{\omega^2}{\omega_n^2} \right)^2 + 4\zeta^2 \left(\frac{\omega}{\omega_n} \right)^2 \right]^{1/2} \quad (9-2-36)$$

For frequencies much less than ω_n Eq. 9-2-25 can be approximated by

$$\text{number of db} \approx 20 \log_{10} 1 = 0 \quad ; \quad \omega \ll \omega_n \quad (9-2-37)$$

For frequencies much greater than ω_n the approximation made is

$$\text{number of db} \approx 20 \log_{10} \left[\left(-\frac{\omega^2}{\omega_n^2} \right)^2 \right]^{1/2} = 40 \log_{10} \left(\frac{\omega}{\omega_n} \right) \quad (9-2-38)$$

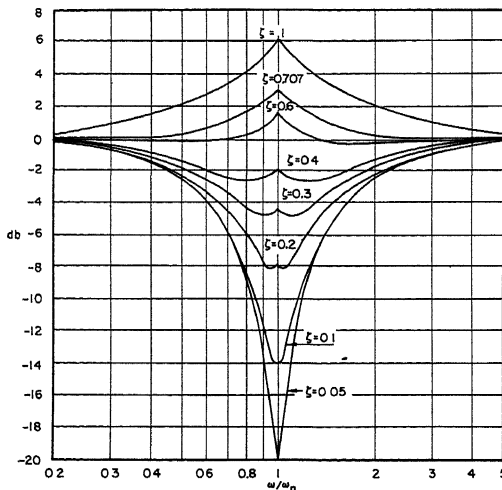


Fig. 9-2-12.

This is plotted as a straight line through $\omega = \omega_n$ with a slope of 40 db/decade. Near ω_n there are generally drastic corrections which depend on the damping ratio ζ . Fig. 9-2-12 shows the general correction curves for various values of ζ . These values must be added to the asymptotic curves.

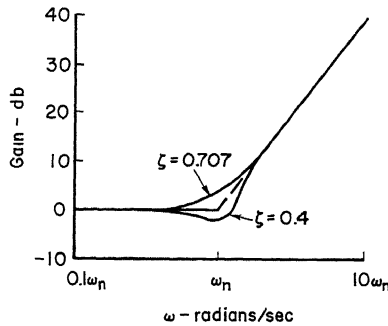


Fig. 9-2-13.

Fig. 9-2-13 illustrates two typical Bode plots for quadratic factors with $\zeta = 0.4$ and $\zeta = 0.707$. If the quadratic factor appears in the denominator of the transfer function, the curves of Fig. 9-2-13 are spun about the zero db line and the break is down. The correction curves of Fig. 9-2-12 must also be read as negative values in this case. The effect of including the gain term K is still to shift the relative curve up or down.

9-3. Bode Diagram Phase Curves. The steady-state transfer function contains not only gain or amplitude information, but also relative phase angle characteristics. Under steady-state sinusoidal excitation the steady-state transfer function gives the phase angle of the output sinusoid with respect to the phase of the input sinusoid. Various expressions for the phase angles of typical transfer function terms follow. For a transfer function such as

$$G(s) = \frac{K(s + z_1)}{s^n(s + p_1)^2(s^2 + 2\zeta\omega_n s + \omega_n^2)} \quad (9-3-1)$$

with s replaced by $j\omega$ the angle of $G(j\omega)$ is

$$\text{ang } G(j\omega) = \tan^{-1}\left(\frac{\omega}{z_1}\right) - n\left(\frac{\pi}{2}\right) - 2 \tan^{-1}\left(\frac{\omega}{p_1}\right) - \tan^{-1}\left[\frac{2\zeta(\omega/\omega_n)}{1 - (\omega/\omega_n)^2}\right] \quad (9-3-2)$$

The over-all angle is composed of the angles of the component terms $j\omega + z_1$, $(j\omega)^{-n}$, $(j\omega + p_1)^{-2}$, and $[(\omega_n^2 - \omega^2) + j2\zeta\omega_n\omega]^{-1}$. There is no dependence whatsoever on the gain constant. The simplest of these terms, $(j\omega)^{-n}$, leads to a constant phase shift of $-n\pi/2$ radians. This phase shift is intimately associated with its magnitude plot, which is an

infinitely long line with a slope of $-20n$ db/decade. This association of $\pi/2$ radians phase shift with each 20 db/decade or 6 db/octave gain curve slope is no mere item of chance. It is part of a broad general relationship between amplitude and phase that holds for a class of transfer functions said to be minimum-phase. These functions have all their poles and zeros in the left-hand half s -plane or on the $j\omega$ -axis. By use of complex variable theory Bode has shown that the phase angle of a minimum-phase transfer function is uniquely replated to the various slopes of its gain curve segments as developed in Art. 9-2. The phase shift at a certain frequency is determined in varying amounts by slopes of all segments of the gain curve. Segments near the frequency at which the phase is of interest have the greatest effect, while more remote segments have a smaller effect. Theoretically, an infinitely long line with a slope of ± 20 db/decade is required for $\pm 90^\circ$ phase shift; however, a segment with ± 20 db/decade slope which extends a decade above and below a frequency of interest gives essentially $\pm 90^\circ$ phase shift. This effect is evident when consider-

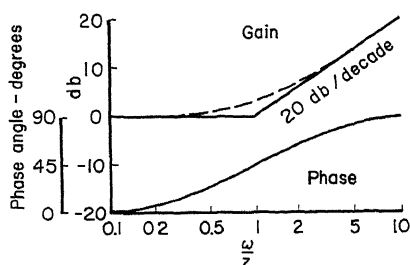


Fig. 9-3-1.

ing a term of the form $j\omega + z$ with an angle $\tan^{-1}(\omega/z)$. Fig. 9-3-1 shows both the phase curve and the relative gain curve for this term. For frequencies much greater than $\omega = z$, the phase angle is essentially 90° , as the effect of the zero slope gain segment below $\omega = z$ is small. At frequencies much less than $\omega = z$ the zero slope gain segment is most effective and causes 0° phase shift. For frequencies near $\omega = z$ both gain curve segments have some effect and the phase curve goes through a transition from 0° to 90° . The phase curve for $\tan^{-1}(\omega/z)$ possesses symmetry about $\omega = z$. The phase shift at a general frequency $\omega = z/a$ is greater than zero by the same amount that the phase shift at $\omega = az$ is less than 90° . Probably the easiest way to sketch the Bode phase curve for a simple zero or pole term is to remember a few values of $\tan^{-1}(\omega/z)$ or to use the tangent scale on a slide rule. While straight-line approximations exist, they are of questionable accuracy. If the phase angle for a pole term is desired, the phase curve becomes lagging or negative instead of positive as for the zero term sketched in Fig. 9-3-1. For a zero or pole of multiple order the phase values above are multiplied by the order of the zero or pole. Addition or subtraction of component angle curves allows

the complete phase curve for a more complex steady-state transfer function to be constructed.

The quadratic transfer function term still remains to be considered. Such terms give the angle $\pm \tan^{-1}[2\zeta(\omega/\omega_n)/(1 - \omega^2/\omega_n^2)]$ from Eq. 9-3-2. From the discussion of gain curve slopes versus phase angles, it should be suspected that this term results in a phase curve that starts at 0° for $\omega \ll \omega_n$ and goes to 180° for $\omega \gg \omega_n$. This is true; however the transition between these two values depends on ζ except at $\omega = \omega_n$, where the phase angle is $\pm 90^\circ$. Fig. 9-3-2 shows a family of curves that allow sketching the quadratic-term phase curves.

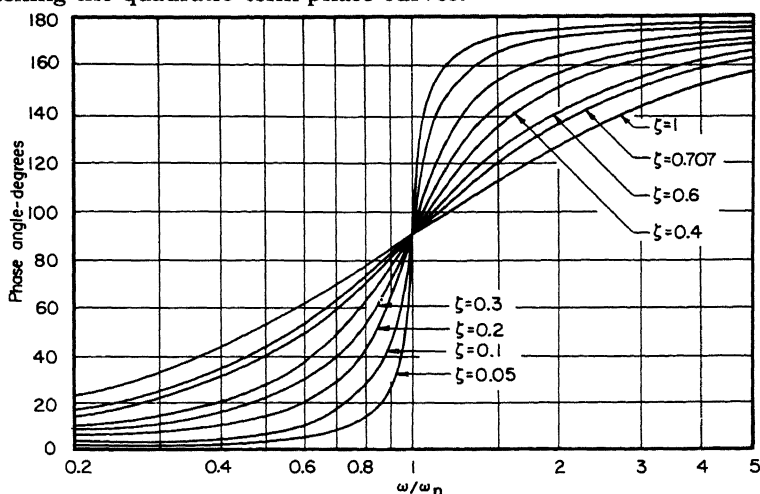


Fig. 9-3-2.

A power-series expansion for the arctangent function is frequently useful in analytical problems which involve finding frequencies for given total phase shift. For a term such as $\tan^{-1}(\omega/z)$ the expansion is

$$\tan^{-1}\left(\frac{\omega}{z}\right) = \left(\frac{\omega}{z}\right) - \frac{1}{3}\left(\frac{\omega}{z}\right)^3 + \frac{1}{5}\left(\frac{\omega}{z}\right)^5 - \cdots ; \quad \omega < z \quad (9-3-3)$$

or

$$\tan^{-1}\left(\frac{\omega}{z}\right) = \frac{\pi}{2} - \left(\frac{z}{\omega}\right) + \frac{1}{3}\left(\frac{z}{\omega}\right)^3 - \frac{1}{5}\left(\frac{z}{\omega}\right)^5 + \cdots ; \quad \omega > z \quad (9-3-4)$$

Rather surprising accuracy can be obtained by considering only the first or linear term in Eqs. 9-3-3 and 9-3-4. The error as a per cent of the

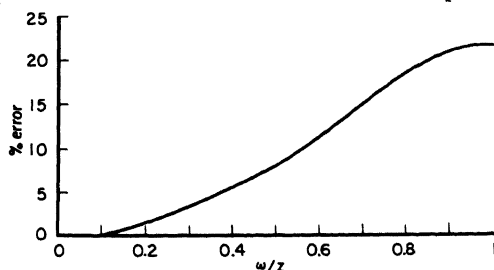


Fig. 9-3-3.

correct value is shown in Fig. 9-3-3. As long as the ratio ω/z is not over about 0.6, reasonable accuracy is obtained from

$$\tan^{-1}\left(\frac{\omega}{z}\right) \approx \frac{\omega}{z} \quad ; \quad \frac{\omega}{z} < 0.6 \quad (9-3-5)$$

or

$$\tan^{-1}\left(\frac{\omega}{z}\right) \approx \frac{\pi}{z_2} - \frac{z}{\omega} \quad ; \quad \frac{z}{\omega} < 0.6 \quad (9-3-6)$$

For the phase angle of a quadratic term the series expansion can be made in a similar manner.

Example 9-3-1. A sketch of the phase curves for Example 9-2-4 is to be made.

$$G(j\omega) = \frac{10(j\omega/6 + 1)}{j\omega/60 + 1} \quad (9-3-7)$$

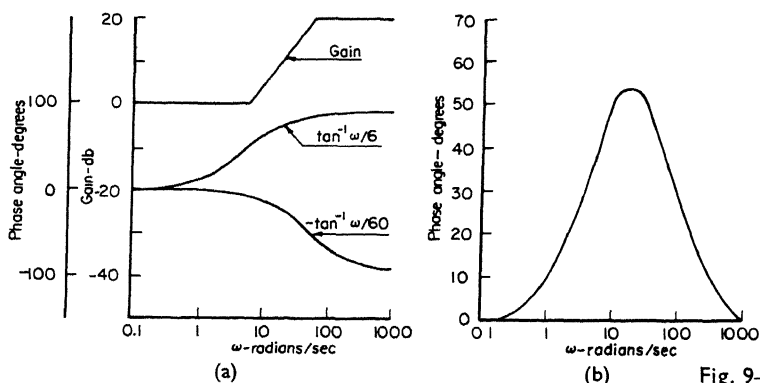


Fig. 9-3-4.

Fig. 9-3-4a shows the asymptotic gain curves and the component phase angle curves due to numerator and denominator terms. Fig. 9-3-4b shows the composite phase curve obtained by adding the component curves of Fig. 9-3-4a.

Example 9-3-2. A Bode phase plot is to be made for

$$G(s) = \frac{100}{s(s^2 + 8s + 100)} \quad (9-3-8)$$

$$\text{ang } G(j\omega) = -\frac{\pi}{2} - \tan^{-1} \frac{(0.8)(\omega/10)}{1 - (\omega/10)^2} \quad (9-3-9)$$

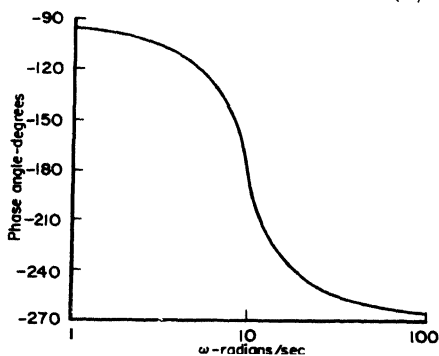


Fig. 9-3-5.

For this case $\zeta = 0.4$. Fig. 9-3-5 shows the exact curve obtained from Fig. 9-3-2 and the asymptotic approximation.

Example 9-3-3. The phase angle for the $G(j\omega)$ used in Example 9-2-5 is to be computed at $\omega = 10$ radians/sec by using the linear terms of the power series for the arctangent:

$$G(j\omega) = \frac{10(j\omega/3 + 1)}{j\omega(j\omega/15 + 1)(j\omega/30 + 1)^2} \quad (9-3-10)$$

$$\text{ang } G(j\omega) = \tan^{-1}\left(\frac{\omega}{3}\right) - \frac{\pi}{2} - \tan^{-1}\left(\frac{\omega}{15}\right) - 2 \tan^{-1}\left(\frac{\omega}{30}\right) \quad (9-3-11)$$

$$\approx \left(\frac{\pi}{2} - \frac{3}{10}\right) - \frac{\pi}{2} - \frac{10}{15} - \frac{20}{30} = -1.63 \text{ radians or } -93.4^\circ \quad (9-3-12)$$

The exact value is -87.2° , which shows reasonable accuracy even though one term, $\omega/15$, has a value of 0.667.

9-4. Stability Requirements on Bode Plots. In previous work with Nyquist diagrams the concepts of gain and phase margin are introduced and related to relative and absolute stability for simple feedback systems. The basic requirement is for a positive phase margin. In other words the angle of phase lag of the open-loop function $G(j\omega)$ must be less than 180° at the frequency at which $|G(j\omega)|$ is 1 or zero decibels. This requirement can be applied directly on the Bode plots to determine whether or not a control system is stable.

A slight alteration of the method of applying the phase margin condition is very useful. Since the phase angle for minimum phase systems is directly related to the slopes of the Bode magnitude plot, an approximate stability criterion can be based on the slopes of the magnitude plot. On such Bode magnitude plots long asymptotic segments with $-20n$ db/decade slopes result in $-n\pi/2$ radians phase shift. Thus if a Bode magnitude plot crosses the zero decibel gain axis with a -40 db/decade sloping segment which starts a few octaves below the crossover frequency and extends a few octaves above crossover, the phase lag is very close to 180° and the system is near instability. Therefore a goal in system design and compensation is to have the Bode magnitude plot of $G(j\omega)$ cross the zero decibel axis with a -20 db/decade slope or with a very short -40 db/decade sloping segment with a -20 db/decade sloping segment on a least one side. This scheme is approximate and quite flexible since the phase shift at any given frequency depends on the slopes of all magnitude plot segments. However the effect of remote segments is weighted much less than that of segments near a frequency of interest.

An additional aid in studying system stability when using frequency response methods is the use of the constant M circles which are introduced earlier. The M contours allow the peak closed-loop magnification, $|C(j\omega)/R(j\omega)|_{\max}$ to be checked while working with an open-loop transfer function. The constant M contours are circles only on the Nyquist plots

When working with Bode diagrams, however, it is convenient to present the M contours on a graph with phase margin as one coordinate and decibels gain as the other coordinate. Such a graph is called a Nichol's

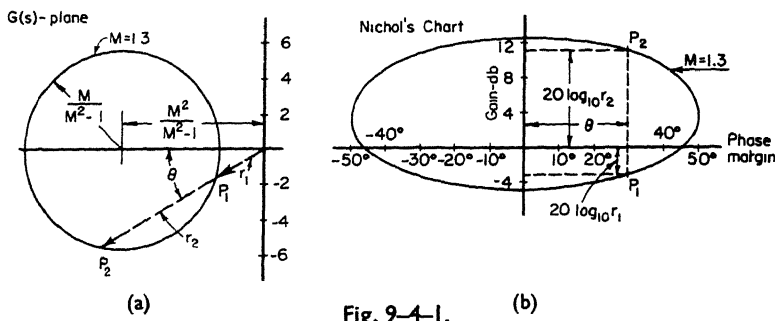


Fig. 9-4-1.

Chart. Fig. 9-4-1a illustrates an M circle in the Nyquist or polar plane and Fig. 9-4-1b shows this M contour on the Nichol's Chart as well as the method of transferring particular points. A transparent sheet with a Nichol's Chart having several M circles is very useful for studying system gain changes. This use is illustrated in later examples. For a given $G(j\omega)$ it is possible to transfer an M contour to the actual Bode plot; however it is less work to transfer the Bode plot information to the Nichol's chart. The following examples clarify some of the above remarks.

9-5. The Gain Adjustment Problem. In nearly every control system certain gain adjustments are necessary. The Bode plot provides a means for studying the phase margin while the gain is varied. This is accomplished by sliding the decibel scale on the Bode plot.

Example 9-5-1. As a first example a unity feedback system with

$$G(s) = \frac{K(s + 10)}{s(s + 4)(s + 20)} \quad (9-5-1)$$

is studied by using the Bode plot. The gain is to be set for a phase margin of 45° . In preparing to plot the Bode diagram, $G(j\omega)$ is rearranged as follows:

$$G(j\omega) = \frac{(K/8)(j\omega/10 + 1)}{(j\omega)(j\omega/4 + 1)(j\omega/20 + 1)} \quad (9-5-2)$$

A convenient step is to let the gain term $K/8$ be unity temporarily while plotting the Bode plot. K is to be adjusted anyway, and this step simplifies plotting. Fig. 9-5-1 shows the Bode plot for $K/8 = 1$ as well as the additional gain required for the zero decibel crossover to occur at a frequency which gives 45° phase margin. The added gain of 22 decibels or 12.6 means the final over-all value of K is $12.6 \times 8 = 101$. In this problem, with no restriction on the gain at some particular frequency or on the velocity constant, a simple gain adjustment is all that is needed. Generally, changes in the system dynamics are required and the shape of the gain and phase curves must be altered. It might be mentioned that the system velocity constant can be found by extending the -20 db/octave sloping segment for $(j\omega)^{-1}$, as set in the final step of the design, until it intersects the zero decibel axis, at which point the

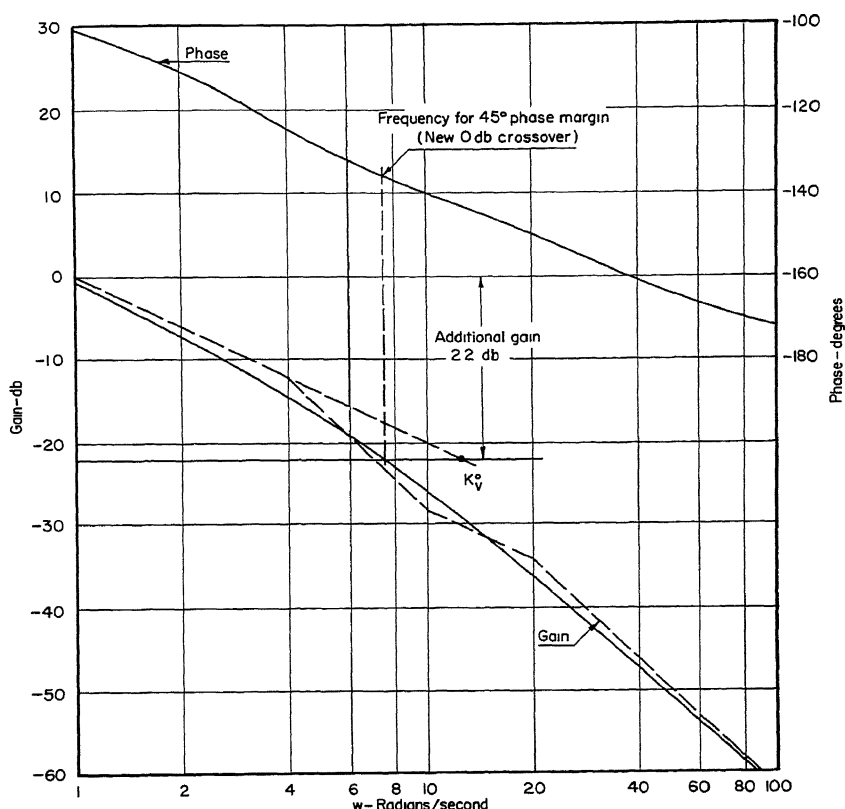


Fig. 9-5-1.

frequency is numerically equal to K_v . This procedure results from the following line of reasoning. The method used in rearranging $G(j\omega)$ to form factors of the form $j\omega/p + 1$ results in isolation of the velocity constant K_v , as in Eq. 9-5-2 where $K_v = K/8$. If only the -20 db/decade sloping segment of $G(j\omega)$ due to $(j\omega)^{-1}$ is considered at very low frequencies,

$$20 \log_{10} |G(j\omega)| \approx 20 \log_{10} K_v - 20 \log_{10} \omega \quad (9-5-3)$$

Thus when $\omega = K_v$, the result is zero decibels. Also the velocity constant in decibels can be read from the gain curve at $\omega = 1$. In this example the final K_v is 12.6.

Example 9-5-2. As a second example a unity feedback system with

$$G(s) = \frac{K}{s(s+1)(s+2)} \quad (9-5-4)$$

is to have K adjusted for an $M = 1.3$ by using the Nichol's chart. The $M = 1.3$ contour is also to be mapped onto the Bode plot. This example is an alternate study of a system used in Example 8-2-1. Fig. 9-5-2 shows the Bode plot for $K = 2$ and part of the $M = 1.3$ contour as discussed below. By reading the values of phase margin and the gain in decibels at a given frequency, the Bode plot information is transferred to the Nichol's chart of Fig. 9-5-3. The $M=1.3$ contour in this figure is only one of a family which can either be transferred from the Nyquist diagram, as indicated in Fig. 9-4-1b, or constructed directly from the relationship between

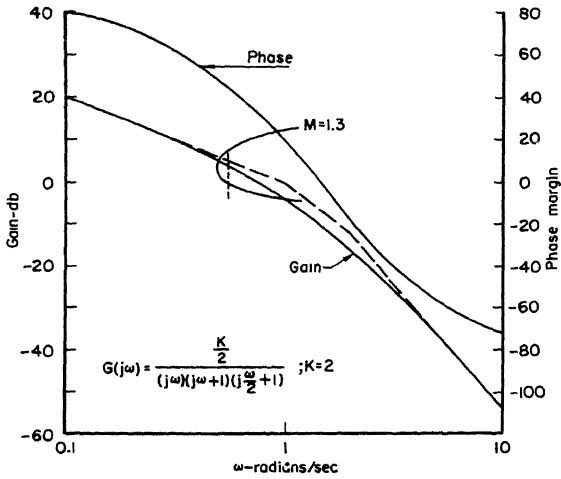


Fig. 9-5-2.

open- and closed-loop transfer functions. A transparent overlay of a Nichol's chart with the M contours on it is useful in adjusting system gain. The overlay can be slid up and down along the decibel scale to show the effects of adjustments in the

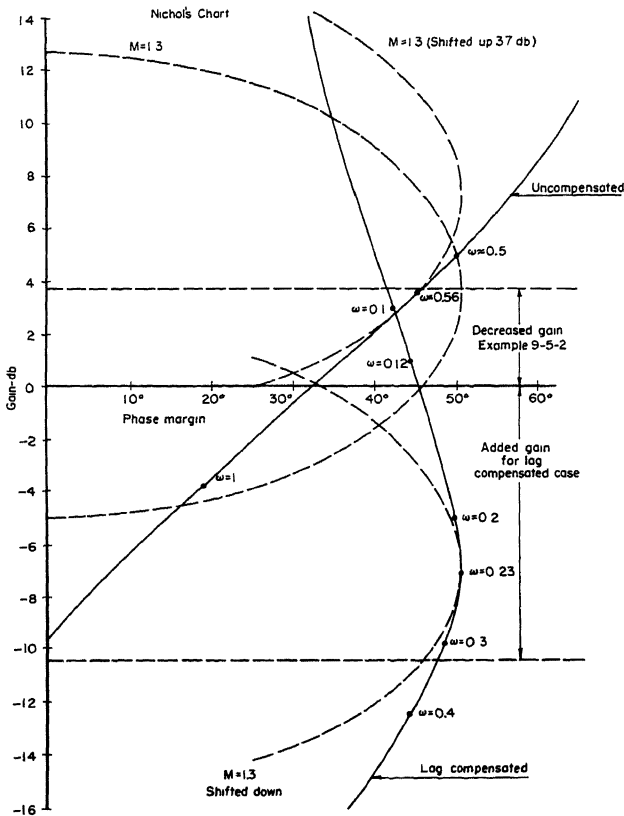


Fig. 9-5-3.

open-loop gain K . In this example the $M = 1.3$ curve has to be slid up about 3.7 db to make it tangent to the system curve with $K = 2$. The tangency occurs at a frequency of about 0.56 radians/sec. This means that the gain K has to be decreased by 3.7 db for $M_m = 1.3$. Since K was assumed to be equal to 2 in plotting the original Bode plot, the final value of K is $(2)(10^{-3.7/20})$ or 1.3. Once a particular Bode phase plot is determined, the M contours can be transferred from the Nichol's chart to the Bode plot as follows. Points with a particular phase margin on an M contour on the Nichol's chart are determined. The frequency at which this particular phase margin occurs is located on the Bode plot and the M circle gain values are then located at this frequency. In this manner the $M = 1.3$ contour of Fig. 9-5-3 is transferred to the Bode plot of Fig. 9-5-2. The gain adjustment can then be accomplished in either plane.

9-6. Lag Compensation on the Bode and Nichol's Diagrams.¹

The requirement of large low-frequency gain or a large K_v for accuracy can sometimes be met by using a lag network as illustrated in Chapter 8. The lag network transfer function

$$G_1(j\omega) = a \frac{j\omega + a/a}{j\omega + a} = \frac{j\omega a + 1}{j\frac{\omega}{a} + 1} \quad ; \quad a < 1 \quad (9-6-1)$$

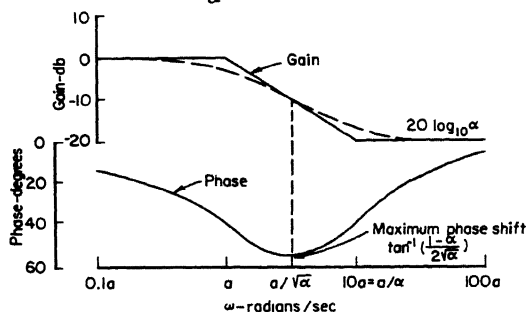


Fig. 9-6-1.

when sketched on the Bode diagram appears as shown in Fig. 9-6-1. Generally the open-loop gain K is increased so as to return the high-frequency region of a compensated Bode plot near its level before compensation. Care must be exercised to prevent the added phase lag from causing instability.

Example 9-6-1. The system of Example 9-5-2 is to be compensated using a lag network with $a = 0.1$ and with the upper break frequency at $\omega = 0.1$. If possible the gain is to be set for $M = 1.3$. Fig. 9-6-2 shows the Bode magnitude and phase curves for $K = 2.0$. It should be pointed out that the lag network break frequencies must be located so that the gain curve has a -20 db/decade segment of appreciable length above the upper break frequency of the lag network. This segment will allow the phase to recover toward $+90^\circ$ phase margin while the gain curve is $20 \log a$ below the uncompensated curve. This is illustrated by the curves of Fig. 9-6-2 for $0.1 < \omega < 1.0$. The gain is set so that crossover occurs near the frequency of maximum phase margin on the -20 db/decade segment. In order to set the gain, the information in Fig. 9-6-2 is transferred to the Nichol's chart of Fig. 9-5-3.

¹ John L. Bower and Peter M. Schultheiss, *Introduction to the Design of Servomechanisms*, John Wiley & Sons, Inc., New York, 1958, pp. 175-182.

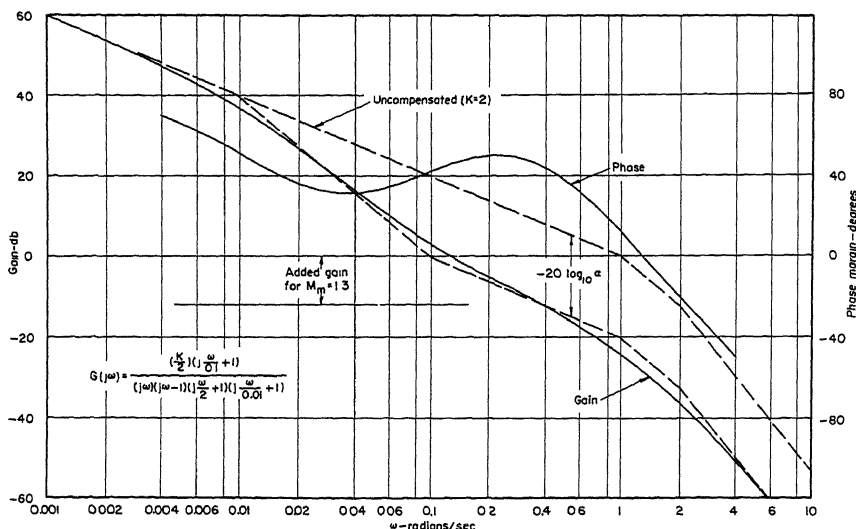


Fig. 9-6-2.

It is apparent from Fig. 9-5-3 that the only hope for placing the lag-compensated system curve tangent to the $M = 1.3$ contour with reasonable values of K is to raise the gain by approximately 10.4 decibels so that the right extremity of the $M = 1.3$ contour touches the system curve at about $\omega = 0.23$ radians/sec. This is found by sliding the M contour overlay down by 10.4 db. This means an increase in gain of $10^{10.4/20} = 3.32$ times. Thus the final system K is (2) ($3.32 = 6.64$) and the system velocity constant is 3.32. In this example the ultimate loss of 20 db gain at high frequencies due to the lag network is not fully recovered by the permissible increase in K . In fact, only about half of this loss is recovered. If the system gain is raised 20 db, a phase margin of only 26° results as compared with 50° for the gain used in the solution. As mentioned above, the phase margin can be used to set the gain if only an approximate M_m is desired. Generally 45° to 50° phase margin gives $M_m = 1.3$. The effect of the lag network on the original phase curve should be noted, as well as the recovery of phase angle in the region of -20 db/decade gain segments. Basically, lag compensation reduces gain before crossover by introducing a phase lag at low frequencies where such lag can be tolerated. There must remain a -20 db/decade segment in the gain curve at crossover.

9-7. Lead Compensation on the Bode and Nichol's Diagrams.

In order to improve the phase margin and stability of a control system a lead network can also sometimes be used. This scheme is introduced in the study of Nyquist diagrams where the effect of leading phase shift is very evident. When plotted on a Bode plot, the characteristics of a passive lead transfer function

$$G_1(j\omega) = \frac{j\omega + b}{j\omega + b/a} = a \frac{j\omega/b + 1}{j\omega a/b + 1} \quad ; \quad a < 1 \quad (9-7-1)$$

appear as shown in Fig. 9-7-1. Thus there is a loss in gain for low frequencies accompanied by a region of phase lead. In terms of the

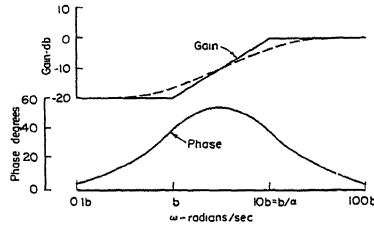


Fig. 9-7-1.

Nichol's chart the low-frequency points are shifted down with those in the region from $\omega = 0.1b$ to $\omega = 10b/a$ shifted down and to the right while the high-frequency ($\omega \gg b/a$) points are not changed.

Example 9-7-1. As a first example of lead compensation using the Bode and Nichol's plots, the system of Example 9-5-2 is reconsidered. This system is to be compensated using a lead network for an $M_m = 1.3$. An inspection of Fig.

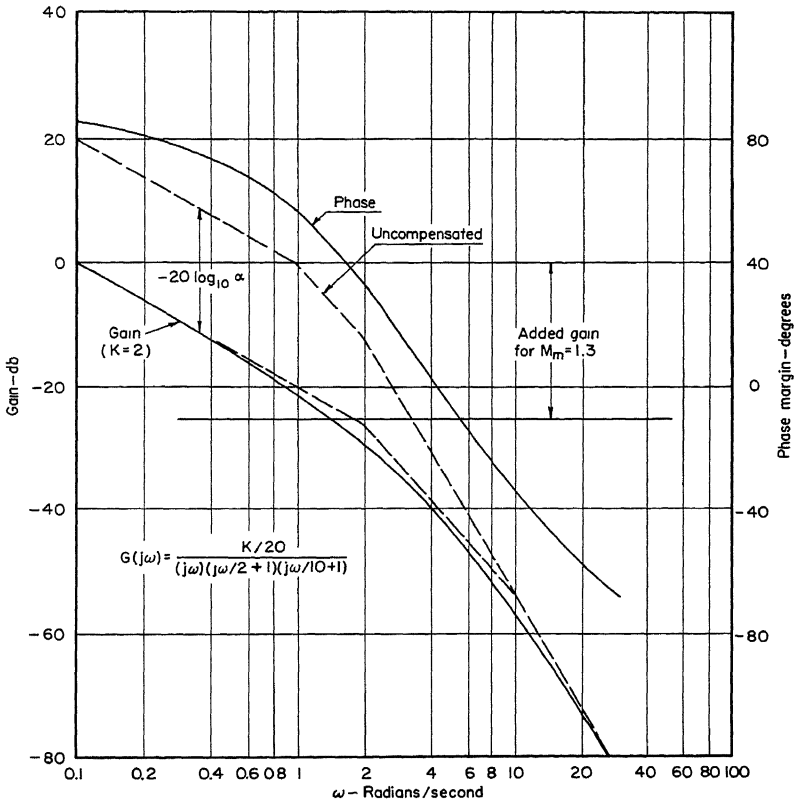


Fig. 9-7-2.

9-5-2 indicates that a maximum phase lead somewhat higher than $\omega = 1.4$ would probably be most effective. The lead function

$$G_1(j\omega) = (0.1) \frac{(j\omega + 1)}{(j\omega/10 + 1)} \quad (9-7-2)$$

is studied. Fig. 9-7-2 shows the Bode plot for the over-all system with $K = 2$ as in Example 9-5-2. The compensated $G(j\omega)$ is

$$G(j\omega) = \frac{K/20}{(j\omega)(j\omega/2 + 1)(j\omega/10 + 1)} \quad (9-7-3)$$

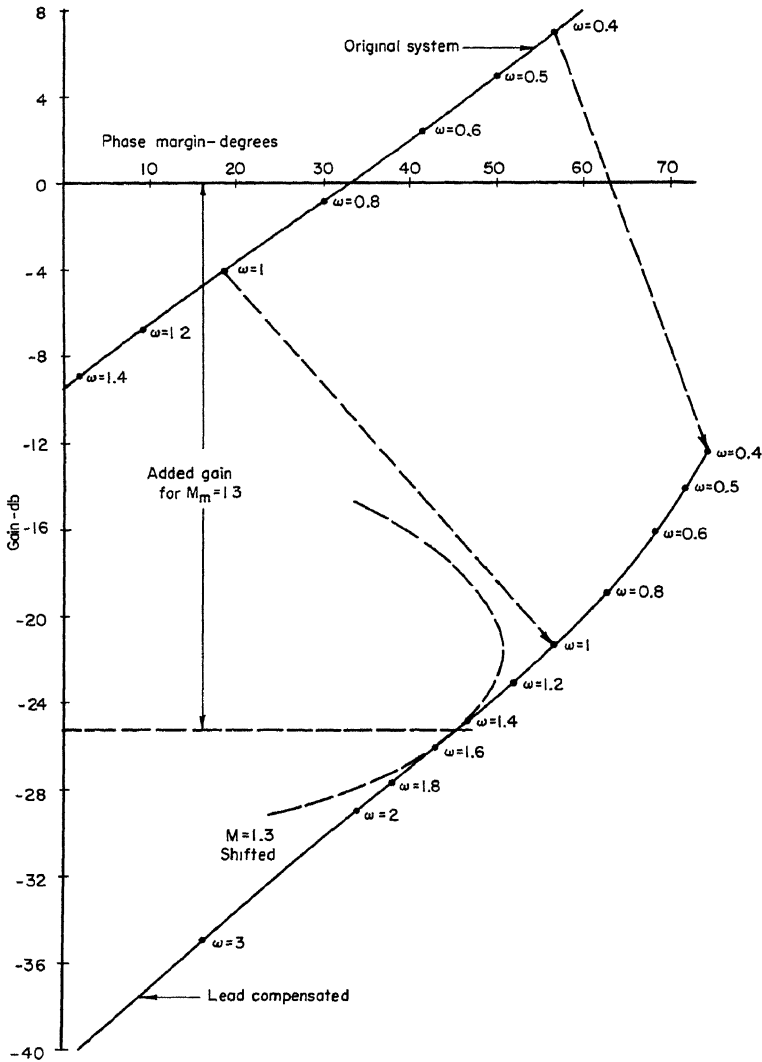


Fig. 9-7-3.

For the same velocity constant as in Example 9-5-2 a gain of 20 db has to be obtained here to offset the loss in low-frequency gain due to the lead network. This would give a phase margin of 60° which is greater than really needed. Some additional gain can be added for $M_m = 1.3$ as determined from the Nichol's chart of Fig. 9-7-3. The actual gain for $M_m = 1.3$ is found to be 25.2 db so that the final K is (2) $(10^{25.2/20})$ or 36.4 and the velocity constant K_v is 1.82. The system resonance is now at $\omega = 1.5$ radians/sec as compared with 0.56 radians/sec without the lead network. Again a phase margin of 45° on the Bode plot would have served as a guide in setting K . If several breaks downward occur close together on a Bode gain plot, lead network compensation is not particularly effective. The use of lead compensation raises the cross-over frequency and, since phase lag increases rapidly in such a case, the added lead phase shift cannot offset the lag to hold a good phase margin.

9-8. A Lag-and-Lead Compensation Template. The simple lag and lead transfer functions of Eqs. 9-6-1 and 9-7-1 have considerable similarity. Except for the multiplier in lead Eq. 9-7-1, the lead and lag equations are reciprocals. It is convenient to plot the magnitude of the compensation functions in decibels versus their phase angles in degrees as shown in Fig. 9-8-1. In fact, the polar symmetry of this figure indicates the possibility of using only the first-quadrant plot and rotating it 180° when changing from lead to lag compensation. The lead transfer function multiplier a has been dropped, and so a gain of $20 \log a$ decibels

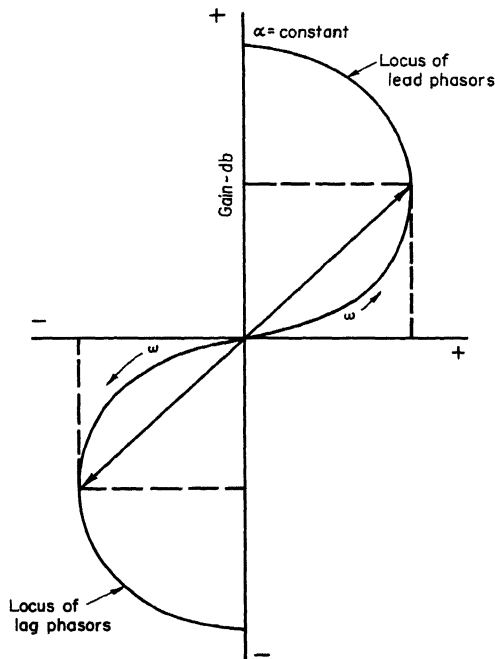


Fig. 9-8-1.

is missing. This actually amounts to restoring the loss in low-frequency gain shown in Fig. 9-7-1 or to raising the whole lead compensation curve above the zero decibel line. Curves of the form shown in Fig. 9-8-1 illustrate how a point on a curve in a Nichol's chart is shifted by lead or lag compensation when the gain factor a is dropped in the lead case. A template of compensation curves based on the plot of Fig. 9-8-1 is shown

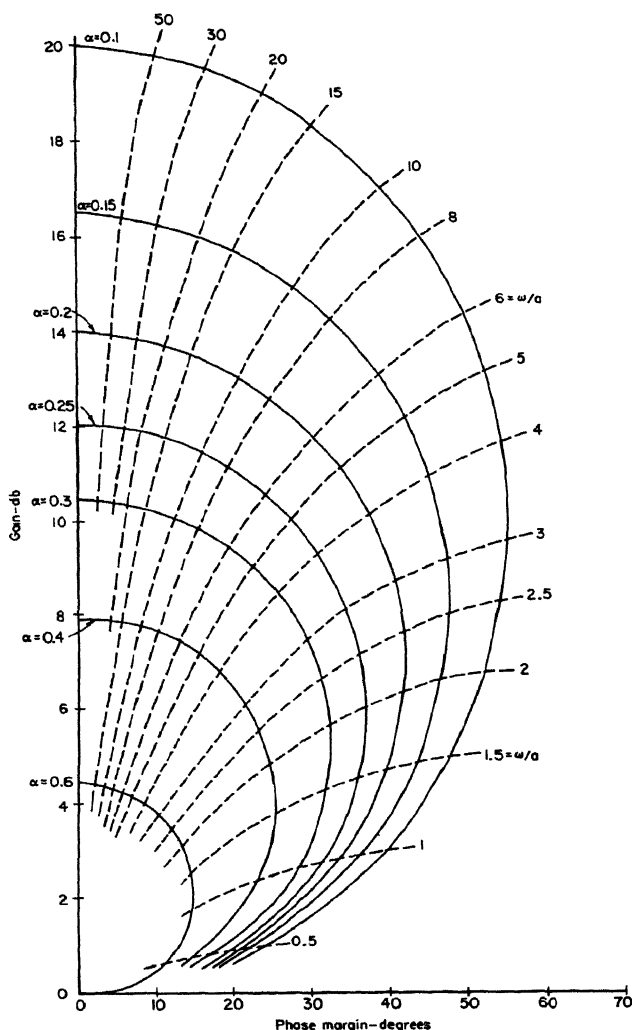


Fig. 9-8-2.

(Adapted with permission from Control Engineering)

in Fig. 9-8-2. This template was suggested by Potts and Thaler.² Curves for fixed values of a and ω/a are included to allow easy determination of the a and a required to shift a point for a given ω to a more desirable

² B. L. Potts and G. J. Thaler, "A Template for Designing Servo Compensators," *Control Engineering* 6 (November, 1959): 139-141.

location. The following examples illustrate the use of the template of Fig. 9-8-2.

Example 9-8-1. Lag compensation of the system of Example 9-5-2 is to be studied using the template of Fig. 9-8-2. An uncompensated system plot with specific values of ω is sketched as shown in Fig. 9-8-3. The template is rotated so

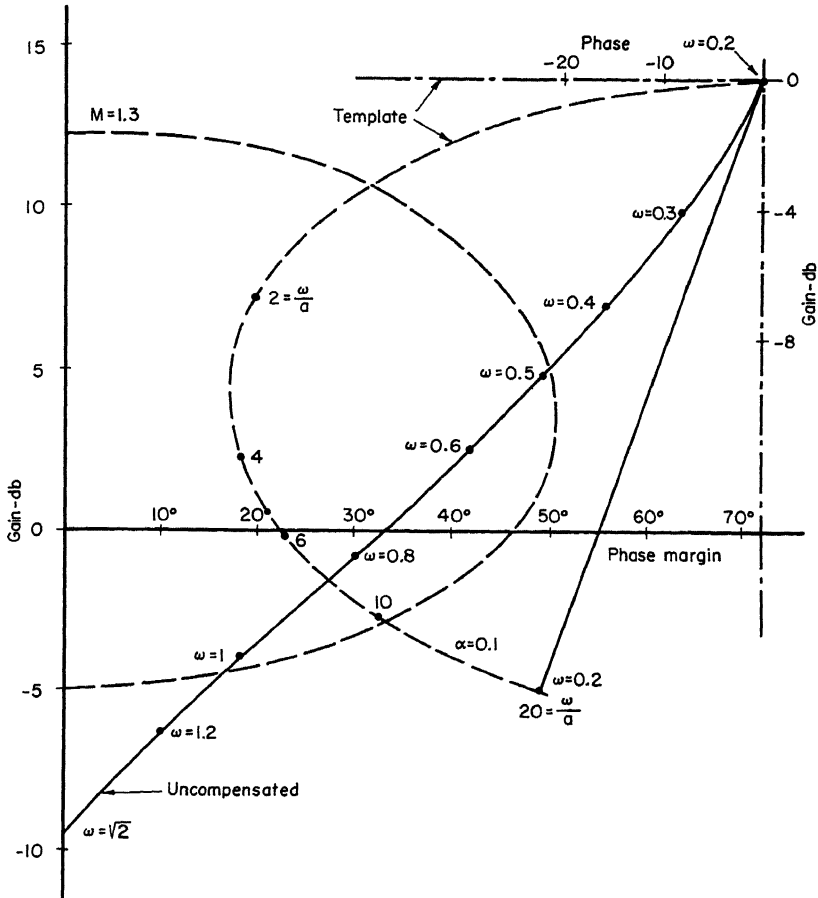


Fig. 9-8-3.

that the origin is at the upper right corner. It is then placed at various points along the uncompensated curve and the lag network shift in such points for various values of a and ω/a is noted. In Fig. 9-8-3 the $a = 0.1$ contour of the template is shown in proper position to study lag network shifts in the point for $\omega = 0.2$. For $\omega/a = 20$ an arrow is included to show the shift. This shift would require an a of 0.01 thereby determining the lag network transfer function. Using these values of a and ω/a , the template can be quickly moved to other points and their shifts determined. Thus rapid study of various lag compensation schemes can be accomplished and a desirable one determined without any recalculations.

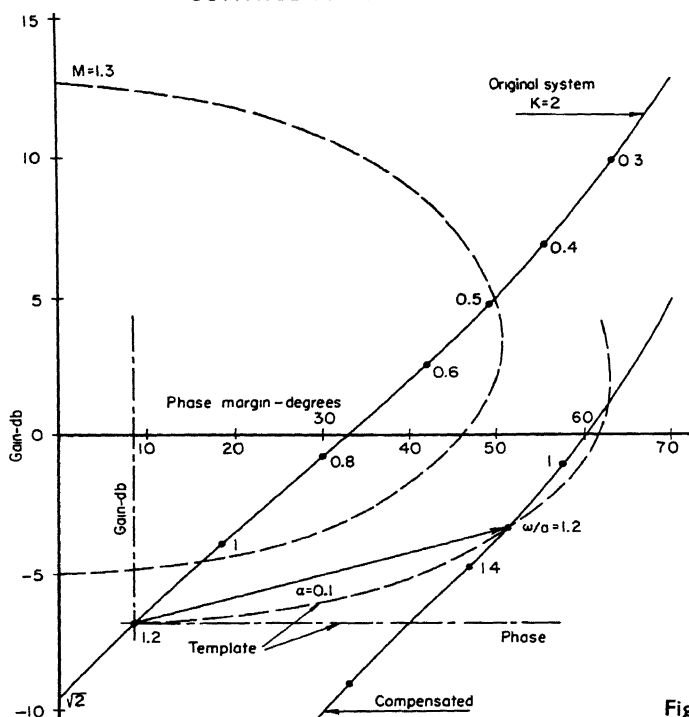


Fig. 9-8-4.

Example 9-8-2. The system of Example 9-5-2 is considered again with lead compensation. Fig. 9-8-4 shows the Nichol's chart plot of the uncompensated system transfer function with $K = 2$, in which case $G(j\omega)$ becomes

$$G(j\omega) = \frac{1}{(j\omega)(j\omega + 1)(j\omega/2 + 1)} \quad (9-8-1)$$

It also includes the $a = 0.1$ contour of the lag-and-lead compensation template in the correct position for studying lead network shifts in the point for $\omega = 1.2$. An arrow illustrates how this point shifts to a new location determined by $\omega/a = 1.2$ or $a = 1$. Since the a gain multiplier is dropped in developing the lead template, the resulting lead-compensated curve is $-20 \log_{10} a$ decibels higher than one which would result from compensation by a passive lead network. This is equivalent to using an increase of this amount in the system gain. Any further increase in system gain which results from the compensation is an actual increase in the system velocity constant. For example, the compensated curve shown in Fig. 9-8-4 for $a = 0.1$ and $a = 1$ can be raised an additional 5.2 decibels for an M_m of 1.3. The actual system gain K is $-20 \log_{10}(0.1) + 5.2$ decibels greater than the K used to plot the original curve or $K = (2)(10^{5.2/20}) = 36.4$. As mentioned above, the velocity constant is 5.2 decibels greater than for the original system (originally $K_v = 2/2 = 1$) or $K_v = 10^{5.2/20} = 1.82$.

The process of studying various other lead compensation schemes is also easily accomplished using the template of Fig. 9-8-2.

9-9. Combined Lag-Lead Compensation on the Bode and Nichol's Diagrams. Combination of the schemes of lag and lead compensation into one compensating network gives composite or lag-lead

compensation. The separate advantages of both schemes when combined lead to considerable increases in the system velocity constant for a given phase margin.

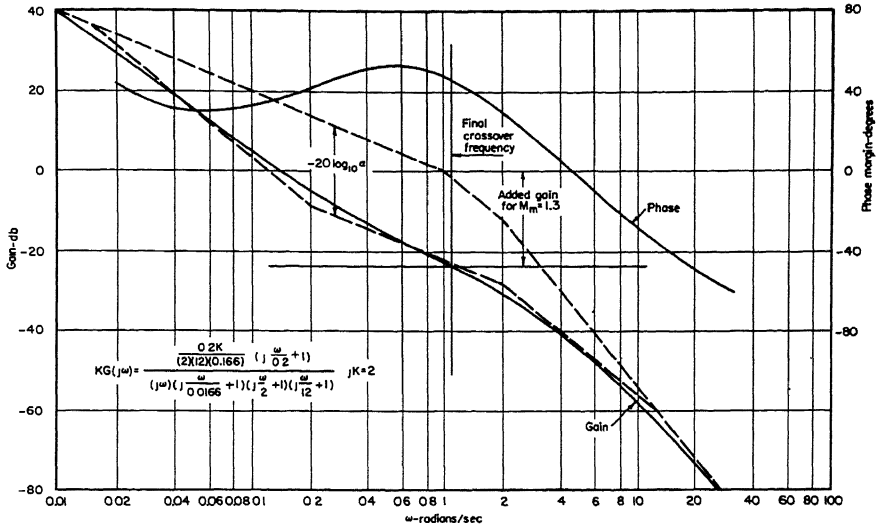


Fig. 9-9-1.

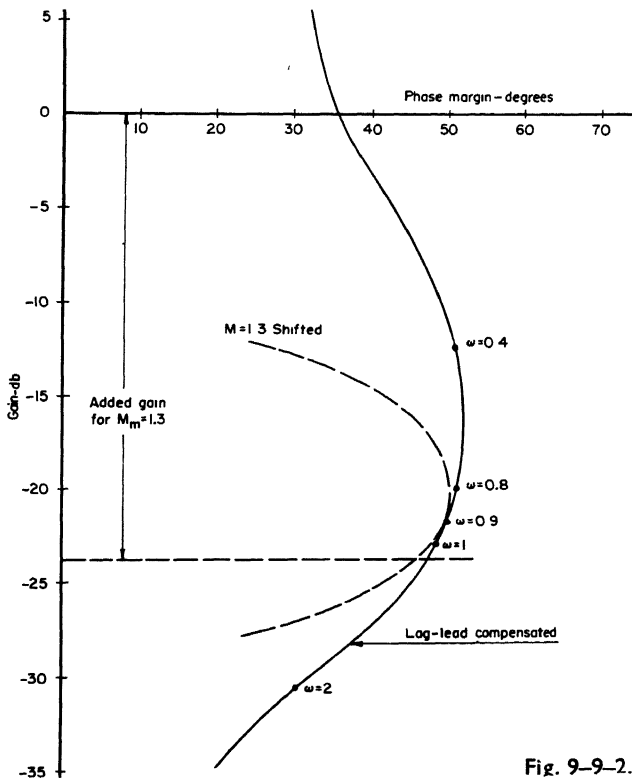


Fig. 9-9-2.

Example 9-9-1. The original system of Example 9-5-2 is to be compensated with a lag-lead equalizer. In order to check Example 8-6-1 the compensation function is taken as

$$G_1(j\omega) = \frac{(j\omega/0.2 + 1)(j\omega + 1)}{(j\omega/0.0166 + 1)(j\omega/12 + 1)} \quad (9-9-1)$$

The lag part is most effective near $\omega = 0.06$ while the lead part reaches its maximum near $\omega = 3.5$. Fig. 9-9-1 shows the compensated Bode plot. The -20 db/decade sloping segment between $\omega = 0.2$ and $\omega = 2.0$ makes it possible to raise the gain considerably while retaining a reasonable phase margin. When the information in Fig. 9-9-1 is plotted on the Nichol's chart of Fig. 9-9-2, an added gain of 23.8 decibels ($K = 31$) is determined for an M_m of 1.3. The final velocity constant is 15.5.

The generally close correspondence between an M_m of 1.3 and a phase margin of 45° allows a rough setting of gain to be made on the Bode plot without using the Nichol's chart. In this example, a 45° phase margin leads to an added gain of 23.6 decibels.

9-10. Conclusion. The ease of sketching Bode plots makes them very useful in many control-system design problems. When used in conjunction with M circle templates on a Nichol's chart plus the lag-and-lead templates of Art. 9-8, the Bode plots are even more useful. They can be extended for use in plotting the frequency response of closed-loop transfer functions. Such plots are useful in determining system bandwidth or the frequency at which the response is 0.707 of its d-c value. With increased use of Bode plots, the student should develop greater facility in determining compensation network and in finding new uses for such plots.

PROBLEMS

9-1. Show that the middle of a frequency interval ω_1 to ω_2 on a log scale is $\sqrt{\omega_1 \omega_2}$. Consider Fig. 9-2-2.

9-2. Construct the Bode magnitude and phase curves for the following transfer functions. Use straight-line magnitude approximations and then correction curves.

(a) $G(s) = 6(s + 1)$

(g) $G(s) = \frac{1}{s(s + 1)(s + 10)}$

(b) $G(j\omega) = \frac{0.6}{j\omega + 1}$

(h) $G(s) = \frac{1000}{s(s^2 + 6s + 100)}$

(c) $G(s) = \frac{28000}{s(s + 10^4)}$

(i) $G(j\omega) = \frac{10^4(j\omega/37.7)^2}{(1 + j\omega/37.7)^2(1 + j\omega/10^6)^2}$

(d) $G(j\omega) = \frac{-100(j\omega/37.7)}{(1 + j\omega/37.7)(1 + j\omega/10^6)}$

(j) $G(s) = \frac{s + 2}{s(s + 0.3)(s + 20)(s + 80)}$

(e) $G(j\omega) = \frac{18000}{(j\omega)(j\omega + 14)(j\omega + 62)}$

(k) $G(s) = \frac{s + 3}{s + 20}$

(f) $G(s) = \frac{100}{s^2 + 6s + 100}$

$$(l) \quad G(s) = \frac{s + 20}{s + 3}$$

$$(n) \quad G(s) = \frac{e^{-0.1s}}{0.1s + 1}$$

$$(m) \quad G(s) = \frac{0.01s^2}{(0.1s + 1)^2}$$

$$(o) \quad G(s) = \frac{-(s - 1)}{s + 1}$$

9-3. Use the method of Art. 9-2 involving ratios to determine the following items for Fig. 9-P-3: (a) $|G_1|$, $|G_3|$; (b) ω_c , $|G_2|$. The slopes are given in exponential units with -1 corresponding to -20 db/decade, etc.

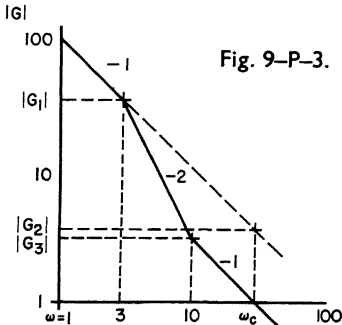


Fig. 9-P-3.

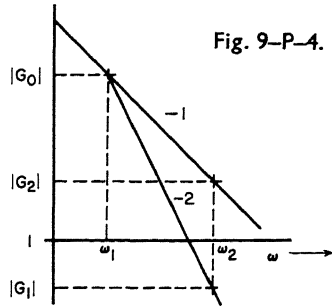


Fig. 9-P-4.

9-4. What must be the ratio of ω_2/ω_1 in the frequency response plot of Fig. 9-P-4 if $|G_1|/|G_2|$ is 10? (Use ratio method.) Slopes are in exponential units.

9-5. Use the arctangent series approximation retaining only linear terms to find the frequency of least phase lag associated with the minimum phase Bode magnitude characteristic of Fig. 9-P-5. Assume the -40 db/decade segments are infinite in extent.

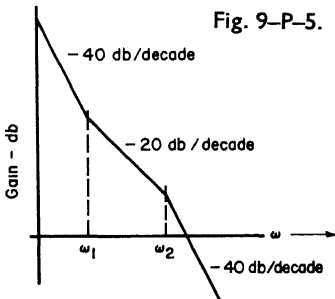


Fig. 9-P-5.

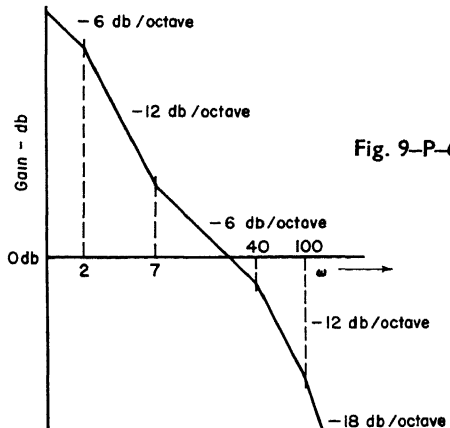


Fig. 9-P-6.

9-6. The Bode magnitude characteristic of Prob. 9-5 is altered as shown in Fig. 9-P-6. Use the result of Prob. 9-5 to set an approximate frequency of least lag on the middle -6 db/octave segment. Determine the phase angle at this frequency by the arctangent series approximation. Plot the actual phase curve in the region $7 < \omega < 40$, and determine the actual least lag and the frequency at which it occurs.

9-7. Repeat Prob. 9-5 except use the Bode magnitude curve of Fig. 9-P-7. Assume that the system from which this characteristic is obtained is minimum-phase.

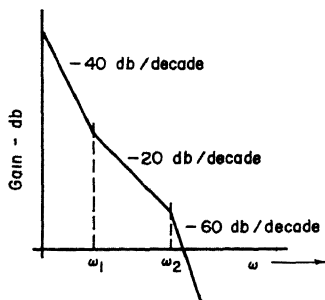


Fig. 9-P-7.

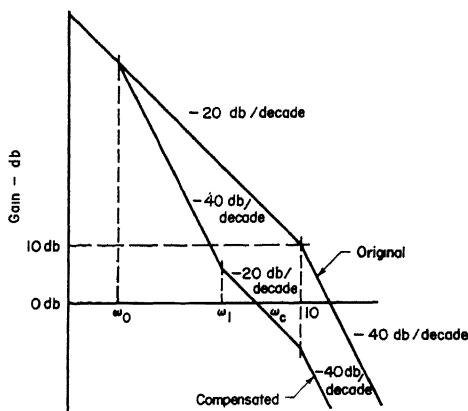


Fig. 9-P-9.

9-8. What minimum ratio ω_2/ω_1 must be used for the system with the Bode magnitude plot of Prob. 9-5 if a phase margin of 50° is desired and crossover is set at the frequency of least lag as determined in Prob. 9-5?

9-9. Problems 9-4, 9-5, and 9-8 build up a basis for approximate lag network design. Use the results of those problems in solving this problem. A sketch of a typical Bode magnitude plot with lag compensation is shown in Fig. 9-P-9. For the values found in Prob. 9-8 determine ω_c , ω_1 , ω_0 , and the difference in gain in the original curve and the compensated curve at frequencies above ω_1 . Neglect the effect of the -20 dB/decade segment at low frequencies when setting ω_c . Is this a reasonable assumption?

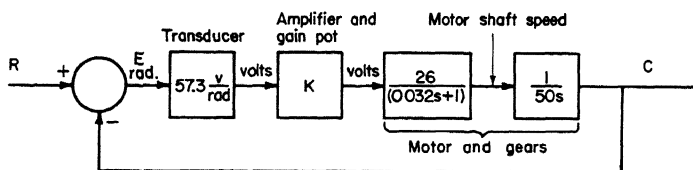


Fig. 9-P-10.

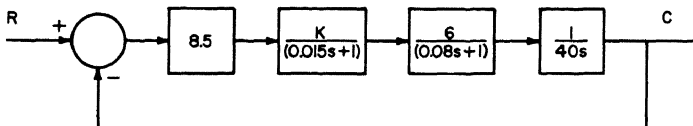


Fig. 9-P-11.

9-10. For the system shown in Fig. 9-P-10 set K for 45° phase margin.

9-11. Set the system gain in Fig. 9-P-11 for $M_m = 1.3$. Plot a Nichol's chart for the system and construct an $M = 1.3$ transparent overlay. (Keep this overlay for future use.) $M = 1.3$ graphs can be found in a number of texts in the servo field. What phase margin results?

9-12. Prove that the frequency of maximum phase lag in the lag compensation function

$$G(j\omega) = \frac{j\omega(aT) + 1}{j\omega T + 1}$$

occurs at $\omega = 1/\sqrt{aT}$ and is given by $\tan^{-1}[(1 - a)/2\sqrt{a}]$.

9-13. Repeat Prob. 9-12 for the lead function

$$G(j\omega) = a \frac{j\omega T + 1}{j\omega aT + 1}$$

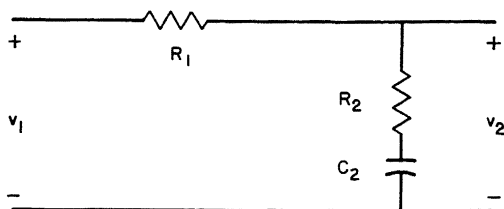


Fig. 9-P-14.

9-14. The system of Prob. 9-11 is to have its velocity constant increased 5 times while holding a phase margin of at least 45° . Design a lag network of the form in Fig. 9-P-14 that can be used ahead of the electronic amplifier and that will accomplish the desired goal. The capacity should not exceed $25\mu\text{f}$ and the resistances 1 megohm.

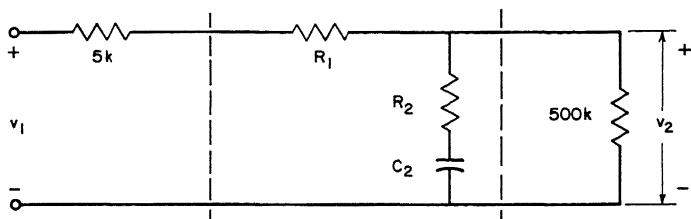


Fig. 9-P-15.

9-15. The design of Prob. 9-14 is complicated by the fact that the error transducer has an output resistance of 6 kilohms and the network is loaded by the amplifier gain pot of 500 kilohms. The configuration of Fig. 9-P-15 is now the circuit. Redesign the network so that the network gives the desired compensation of Prob. 9-14. Why should R_1 be kept reasonably small?

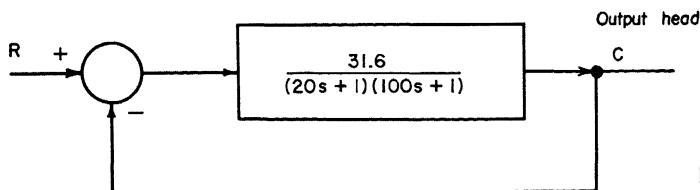


Fig. 9-P-16.

9-16. A certain liquid level regulator system has the block diagram shown in Fig. 9-P-16. An attempt is made to widen the system BW by lead compensation and thereby speed up the response. Add lead compensation to make the BW roughly 0.4 radians/sec and keep the 31.6 position constant. (Use the approximation that

$$\frac{C(j\omega)}{R(j\omega)} = \frac{G(j\omega)}{1 + G(j\omega)} \approx 1 \quad ; \quad |G(j\omega)| \gg 1$$

$$\approx G(j\omega) \quad ; \quad |G(j\omega)| \ll 1$$

to set rough closed loop BW .) Plot the correct $[C/R](j\omega)$ for the original system and for your compensated system. What is the actual bandwidth? How much can the position constant be increased and a 45° phase margin obtained in the compensated case?

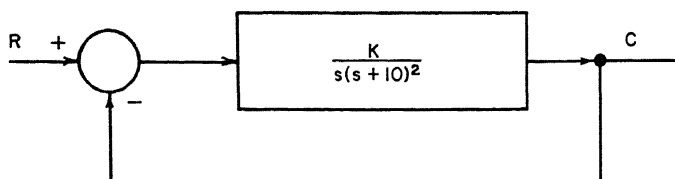


Fig. 9-P-17.

9-17. What value of K gives 45° phase margin for the system of Fig. 9-P-17? Try lead compensation to increase K while maintaining a 45° phase margin. In particular, use the compensation function

$$G_1(j\omega) = (10) \frac{(j\omega + 5)}{(j\omega + 50)}$$

What is the new K and the new velocity constant? Determine the slope of the phase curve of the original $G(s)$ at $\omega = 10$ and of the compensated $G(s)$ at $\omega = 10$. Is lead compensation very effective if the phase lag increases rapidly with frequency? Why?

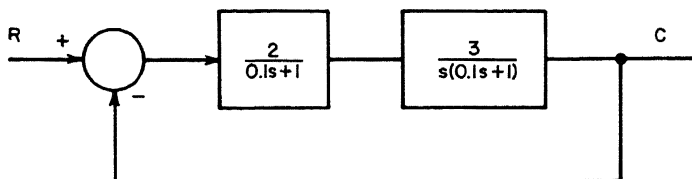
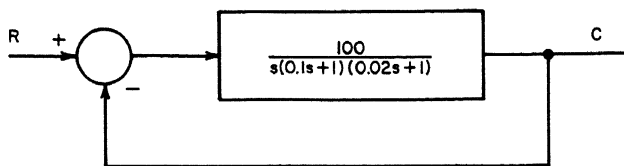
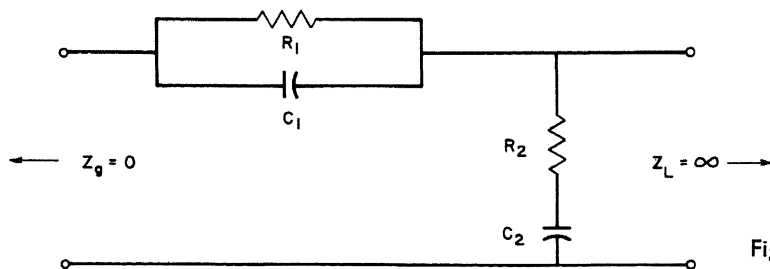


Fig. 9-P-18.

9-18. Construct a transparent overlay of the $a = 0.2$ and the $a = 0.1$ curves from the template of Fig. 9-8-2. Use this template to determine a lag compensation function that decreases steady-state errors to rate of change inputs by a factor of at least 2 for the system of Fig. 9-P-18 while keeping $M_m \leq 1.3$.



(a)



(b)

Fig. 9-P-19.

9-19. Use lag-lead compensation on the system of Fig. 9-P-19a to obtain a 45° phase margin with no change in the velocity constant. An electronic amplifier exists in the system so that an electrical network of the form of Fig. 9-P-19b can be used. Determine suitable values for the parameters for your lag-lead scheme.

10

TRANSIENT AND STEADY-STATE RELATIONSHIPS

10-1. Introduction. Chapter 6 discusses the disadvantages of attempting to use transient methods for design of a control system described by a high order differential equation. For the most part, the disadvantages result from the devious relationship between the coefficients and the zeros of the characteristic polynomial.

Therefore, the sinusoidal frequency response method was developed as an aid in design. The graphical procedures used in either the Nyquist diagram or Bode plots yield considerable information about absolute and relative stability. In addition, these graphical procedures demonstrate the effects of adding compensation to the system. However, the frequency response methods of design also suffer limitations. For example, a system is to have a certain transient response but the designer decides to use sinusoidal frequency response as the basic method of design. The designer interprets the transient specifications into frequency specifications such that the closed-loop system response curve should have a maximum magnitude of $M_m = 1.3$. It can be seen intuitively that two systems with the same M_m can have radically different transient responses as

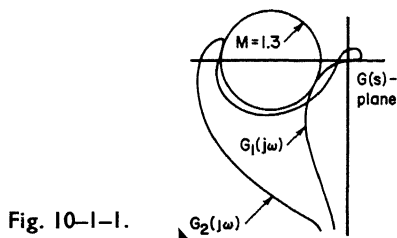


Fig. 10-1-1.

suggested by the $G_1(j\omega)$ and $G_2(j\omega)$ plots of Fig. 10-1-1. Therefore, after the design is completed the transient response must be calculated.

A number of methods have been proposed for correlating the frequency and transient response of systems. Many of these methods are based on the Fourier transforms which are discussed in the next article.

10-2. Fourier Transforms. For notational convenience, the transfer function of the closed-loop system is denoted by $T(s)$ as

$$T(s) = \frac{C(s)}{R(s)} = \frac{G(s)}{1 + H(s)G(s)} \quad (10-2-1)$$

The sinusoidal steady-state response is found in Chapters 3 and 6 (among other chapters) by replacing $C(s)$ and $R(s)$ by their phasor transforms and by substituting $s = j\omega$ in $T(s)$ as

$$T(j\omega) = \frac{C(j\omega)}{R(j\omega)} \quad (10-2-2)$$

It is instructive to retrace this development to see how $T(j\omega)$ can be found more directly.

In Chapter 2, the direct Laplace transform is defined as

$$F(s) = \int_0^{\infty} f(t)e^{-st} dt \quad (10-2-3)$$

In terms of $t(t)$ this becomes

$$T(s) = \int_0^{\infty} t(t)e^{-st} dt \quad (10-2-4)$$

where $t(t)$ is the impulse response of the system. This integral converges if σ is larger than σ_0 , the abscissa of convergence.

After $T(s)$ is obtained, s is replaced by $j\omega$, if the pertinent $\sigma_0 < 0$, to obtain $T(j\omega)$. This substitution implies that the s in the integral of Eq. 10-2-4 is replaced by $j\omega$ as

$$T(j\omega) = \int_0^{\infty} t(t)e^{-j\omega t} dt \quad (10-2-5)$$

Since $t(t)$ is the impulse response of the system with the impulse applied at $t = 0$, $t(t)$ is zero for negative time and the integral can be written as

$$T(j\omega) = \int_{-\infty}^{+\infty} t(t)e^{-j\omega t} dt \quad (10-2-6)$$

This equation is called the direct Fourier transform of $t(t)$ and for the purposes of this book, Eqs. 10-2-5 and 10-2-6 are considered interchangeable.

The direct Fourier transform is essentially the direct Laplace transform with the limitation that $s = \sigma + j\omega$ is restricted to $s = j\omega$. This restriction on σ means that the integral of the direct Fourier transform converges to yield a $T(j\omega)$ only if σ_0 of the corresponding direct Laplace transform is less than zero. Therefore, a function such as e^{3t} has a Laplace transform

$$T(s) = \frac{1}{(s-3)} \quad (10-2-7)$$

whereas the integral of the direct Fourier transform does not converge.

It is possible to derive a Laplace transform such as Eq. 10-2-7 and to formally replace s by $j\omega$ to obtain $T(j\omega)$ as

$$T(j\omega) = \frac{1}{(j\omega - 3)} \quad (10-2-8)$$

This $T(j\omega)$ is not even a Fourier transform because the direct integral as given by Eq. 10-2-6 does not converge for the function ϵ^{3t} . To clarify the notation, it might be desirable to use another symbol to replace $T(j\omega)$ in such a situation. However, the designer that uses the technique of replacing s in the Laplace transform by $j\omega$ to obtain the Fourier transform, might inadvertently make this substitution when it is mathematically improper to do so. The following example is presented to examine the consequences of making such a substitution. A time function $r(t) = \sin 3t$ is applied to a system and the Laplace transform of the output is

$$C(s) = R(s)T(s) = \frac{3}{(s^2 + 3^2)(s - 3)} \quad (10-2-9)$$

The inverse transform $c(t)$ is

$$c(t) = 0.236 \sin(3t - 135^\circ) + 0.167\epsilon^{3t} \quad (10-2-10)$$

In terms of the frequency response methods as used in Chapters 3 and 6, $r(t) = 1 \sin 3t$ is represented by a phasor such as

$$R(j\omega) = 1/\underline{0^\circ} \quad (10-2-11)$$

and $T(j\omega)$ becomes

$$T(j\omega) = \frac{1}{(j3 - 3)} = 0.236/\underline{-135^\circ} \quad (10-2-12)$$

The output as a phasor is

$$C(j\omega) = R(j\omega)T(j\omega) = 0.236/\underline{-135^\circ} \quad (10-2-13)$$

The time function represented by this phasor should yield the output function in steady state

$$c(t)_{ss} = 0.236 \sin(3t - 135^\circ) \quad (10-2-14)$$

When this equation is compared with Eq. 10-2-10 it yields the correct steady-state component of the solution but since the transient component goes off to infinity the system response never settles into this steady-state condition.

In control system work the designer does not intend to produce an unstable system. Therefore, in this practical situation, the limitations upon replacing s in the Laplace transform by $j\omega$ to obtain the Fourier transform, which were discussed above, can be ignored.

Although the inverse Laplace transform integral has not been featured in this book, it is introduced in Chapter 2 as

$$\mathcal{L}^{-1}[T(s)] = \frac{1}{2\pi j} \int_{c-j\infty}^{c+j\infty} T(s)\epsilon^{st} ds = t(t) \quad (10-2-15)$$

The path of integration indicated by this equation is a path parallel to and c units away from the $j\omega$ -axis. The quantity c is chosen so that the path of integration is to the right of all the poles in the complex s -plane.

The inverse Fourier transform can be obtained from the inverse Laplace transform by replacing s by $j\omega$ when such is valid. This requires that the paths of integration be along the $j\omega$ -axis. The inverse Fourier transform is

$$t(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} T(j\omega) e^{j\omega t} d\omega \quad (10-2-16)$$

If the system is stable, i.e. if all poles of the characteristic polynomial are in the left half s -plane, the direct Fourier transform allows the designer to go from the system's impulse response $t(t)$, to its steady-state sinusoidal response $T(j\omega)$, while the inverse Fourier transform allows him to go from the steady-state sinusoidal response to the system's impulse response.

10-3. Convolution. As mentioned in Chapter 2 the multiplication of two Laplace transforms such as $F_1(s)F_2(s)$ gives the transform of a function which can be obtained from what is called the convolution of the two corresponding time functions.

The two time functions are $f_1(t)$ and $f_2(t)$ and their transforms are

$$\begin{aligned} \mathcal{L}[f_1(t)] &= F_1(s) \\ \mathcal{L}[f_2(t)] &= F_2(s) \end{aligned} \quad (10-3-1)$$

The inverse transform of $F_1(s)F_2(s)$ is given by

$$\mathcal{L}^{-1}[F_1(s)F_2(s)] = \int_0^t f_1(t - \tau) f_2(\tau) d\tau = f_1(t) * f_2(t) \quad (10-3-2)$$

or by

$$\mathcal{L}^{-1}[F_1(s)F_2(s)] = \int_0^t f_2(t - \tau) f_1(\tau) d\tau = f_2(t) * f_1(t) \quad (10-3-3)$$

For notational convenience, the product of $F_1(s)F_2(s)$ is denoted by $F_3(s)$ and the inverse transform of $F_3(s)$ is denoted $f_3(t)$ as given by

$$f_3(t) = \mathcal{L}^{-1}[F_3(s)] = \mathcal{L}^{-1}[F_1(s)F_2(s)] \quad (10-3-4)$$

An understanding of the convolution integral is best developed by studying a few simple examples. In each example, $f_3(t)$ is found both as $\mathcal{L}^{-1}[F_3(s)]$ and by using the convolution integral, so that the results may be compared.

Example 10-3-1. For this example both $f_1(t)$ and $f_2(t)$ are the unit-step function as given by

$$f_1(t) = u_{-1}(t) \quad (10-3-5)$$

$$f_2(t) = u_{-1}(t) \quad (10-3-6)$$

The transforms are

$$F_1(s) = \frac{1}{s} \quad (10-3-7)$$

$$F_2(s) = \frac{1}{s} \quad (10-3-8)$$

and
$$F_3(s) = \frac{1}{s^2} \quad (10-3-9)$$

In an example as simple as this, $f_3(t)$ can be found most easily from the inverse transform of Eq. 10-3-9 as

$$f_3(t) = \mathcal{L}^{-1}\left[\frac{1}{s^2}\right] = t \quad (10-3-10)$$

This $f_3(t)$ is now determined by using the convolution integral. In the interval of integration implied by the convolution integral, the unit-step function is equal to 1.

Therefore
$$f_1(t - \tau) = 1 \quad (10-3-11)$$

and
$$f_2(\tau) = 1 \quad (10-3-12)$$

The convolution integral yields

$$\int_0^t f_1(t - \tau)f_2(\tau)d\tau = \int_0^t (1)(1) d\tau = \tau \Big|_0^t = t \quad (10-3-13)$$

Example 10-3-2. For this example both $f_1(t)$ and $f_2(t)$ are the unit-ramp function as given by

$$f_1(t) = u_{-2}(t) \quad (10-3-14)$$

$$f_2(t) = u_{-2}(t) \quad (10-3-15)$$

The transforms are

$$F_1(s) = \frac{1}{s^2} \quad (10-3-16)$$

$$F_2(s) = \frac{1}{s^2} \quad (10-3-17)$$

and
$$F_3(s) = \frac{1}{s^4} \quad (10-3-18)$$

The time function $f_3(t)$ can be found as the inverse transform of Eq. 10-3-18

$$f_3(t) = \mathcal{L}^{-1}\left[\frac{1}{s^4}\right] = \frac{t^3}{6} \quad (10-3-19)$$

This $f_3(t)$ is also determined by using the convolution integral with

$$f_1(t - \tau) = t - \tau \quad (10-3-20)$$

$$f_2(\tau) = \tau \quad (10-3-21)$$

and the integral yields

$$\int_0^t f_1(t-\tau)f_2(\tau)d\tau = \int_0^t (t-\tau)(\tau)d\tau = \left. \frac{t\tau^2}{2} \right|_0^t - \left. \frac{\tau^3}{3} \right|_0^t = \frac{t^3}{6} \quad (10-3-22)$$

Example 10-3-3. Ex. 10-3-2 is reworked using a different approach. $F_3(s)$, the product transform of $F_1(s)F_2(s)$, is unchanged if $F_1(s)$ is multiplied by s and $F_2(s)$ is divided by s as

$$f_3(t) = \mathcal{L}^{-1}[F_3(s)] = \mathcal{L}^{-1}\left[sF_1(s)\frac{F_2(s)}{s}\right] \quad (10-3-23)$$

In accordance with Eq. 10-3-23 and provided the condition noted below is satisfied, the convolution integral can be put into either of the forms

$$f_3(t) = \int_0^t f_1'(t-\tau)f_2^{-1}(\tau)d\tau \quad (10-3-24)$$

or

$$f_3(t) = \int_0^t f_1'(\tau)f_2^{-1}(t-\tau)d\tau \quad (10-3-25)$$

The role of $f_1(t)$ and $f_2(t)$ can be exchanged in these equations. In order for these forms to be correct, the differentiation of the function must be taken in the general sense as discussed in Chapter 2. That is, if $f_1(t)$ has a discontinuity at $t = 0$, $f_1'(t)$ contains an impulse at $t = 0$ in addition to the derivative in the conventional sense. In the previous example

$$f_1(t) = u_{-2}(t) \quad (10-3-26)$$

therefore

$$f_1'(t) = u_{-1}(t) \quad (10-3-27)$$

and

$$f_2(t) = u_{-2}(t) \quad (10-3-28)$$

therefore

$$f_2^{-1}(t) = \int_0^t t dt = \frac{t^2}{2} \quad (10-3-29)$$

The convolution integral yields

$$\int_0^t f_1'(t-\tau)f_2^{-1}(\tau)d\tau = \int_0^t 1 \cdot \frac{\tau^2}{2} d\tau = \frac{t^3}{6} \quad (10-3-30)$$

Example 10-3-4. The $f_1(t)$ and $f_2(t)$ for the next example are chosen as

$$f_1(t) = u_{-2}(t) \quad (10-3-31)$$

$$f_2(t) = \varepsilon^{-at}u_{-1}(t) \quad (10-3-32)$$

the transforms are

$$F_1(s) = \frac{1}{s^2} \quad (10-3-33)$$

$$F_2(s) = \frac{1}{(s+a)} \quad (10-3-34)$$

and

$$F_3(s) = \frac{1}{s^2(s+a)} \quad (10-3-35)$$

The function of $f_3(t)$ can be found by conventional methods as

$$f_3(t) = \mathcal{L}^{-1} \left[\frac{1}{s^2(s+a)} \right] = \frac{1}{a}t - \frac{1}{a^2} + \frac{1}{a^2}e^{-at} \quad (10-3-36)$$

If the convolution integral in the form

$$\int_0^t f_1(t-\tau)f_2(\tau)d\tau \quad (10-3-37)$$

is used, the term

$$\int_0^t (t-\tau)e^{-a\tau}d\tau \quad (10-3-38)$$

is encountered. Although this integral is not difficult to evaluate, it can be avoided completely by using the convolution integral in the form

$$\int_0^t f_1'(\tau)f_2^{-1}(t-\tau)d\tau \quad (10-3-39)$$

In this case

$$f_1'(t) = u_{-1}(t) \quad (10-3-40)$$

and

$$f_2^{-1}(t) = \int_0^t e^{-a\tau}d\tau = \frac{1}{a}(1 - e^{-at}) \quad (10-3-41)$$

Eq. 10-3-39 can now be written as

$$\int_0^t f_1'(\tau)f_2^{-1}(t-\tau)d\tau = \int_0^t 1 \cdot \frac{1}{a}(1 - e^{-a\tau})d\tau = \frac{1}{a}t - \frac{1}{a^2} + \frac{1}{a^2}e^{-at} \quad (10-3-42)$$

10-4. Graphical Interpretation of the Convolution Integral.

Two basic forms have been given for the convolution integral as

$$f_1(t)*f_2(t) = \int_0^t f_1(t-\tau)f_2(\tau)d\tau \quad (10-4-1)$$

or

$$f_2(t)*f_1(t) = \int_0^t f_2(t-\tau)f_1(\tau)d\tau \quad (10-4-2)$$

In some situations, the two time functions may be given graphically, in which case graphical interpretation of the convolution integral is very helpful. For example, $f_1(t)$ and $f_2(t)$ are shown in Fig. 10-4-1a and b respectively. On these sketches, the change of variable from t to τ has already been made. The form of the convolution integral given by Eq. 10-4-1 is the one to be used. As a start in this direction, $f_1(-\tau)$ is sketched in Fig. 10-4-1c and for a specific value of t , $f_1(t-\tau)$ is shown in Fig. 10-4-1d. The product of $f_1(t-\tau)f_2(\tau)$ is shown in Fig. 10-4-1e. The area under the product curve is the integral $\int_0^t f_1(t-\tau)f_2(\tau)d\tau$ for this

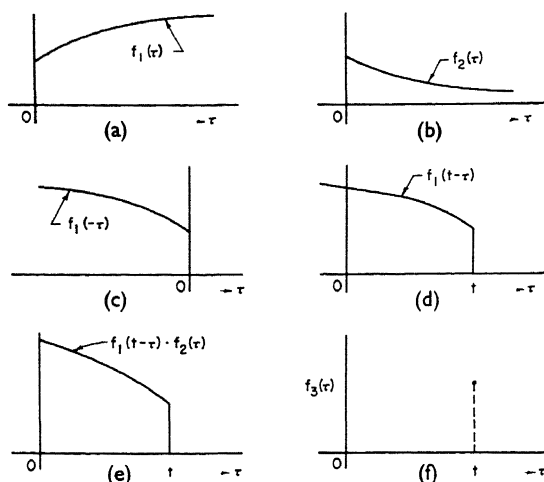


Fig. 10-4-1.

specific value of t and thus one point of the $f_3(t)$ function is obtained as shown in Fig. 10-4-1f.

For each value of t , a different product curve is formed. For each product, the area is calculated and a point on the $f_3(t)$ curve is determined. The $f_1(\tau)$ curve is folded over, so to speak, and then as time t increases, the $f_1(t - \tau)$ curve slides across the figure from left to right.

The other form of the convolution integral as given by Eq. 10-4-2 can be obtained by folding the $f_2(\tau)$ curve and the resulting function $f_2(t - \tau)$ slides across the figure forming the product $f_2(t - \tau)f_1(\tau)$. Since the area under the product curve formed by $f_1(t - \tau)f_2(\tau)$ or that formed by $f_2(t - \tau)f_1(\tau)$ is the same for a given value of t , either form of the convolution integral leads to the same results.

Example 10-4-1. For this example, both $f_1(t)$ and $f_2(t)$ are the unit-step function. In this example, $f_2(t)$ is the function that is folded. Fig. 10-4-2a shows $f_1(\tau)$, Fig. 10-4-2b shows $f_2(t - \tau)$ and Fig. 10-4-2c shows the product curve $f_2(t - \tau)f_1(\tau)$. Since the product curve has a unit height and a base of t , the area under this curve is equal to t and thus the $f_3(t)$ curve shown in Fig. 10-4-2d is obtained.

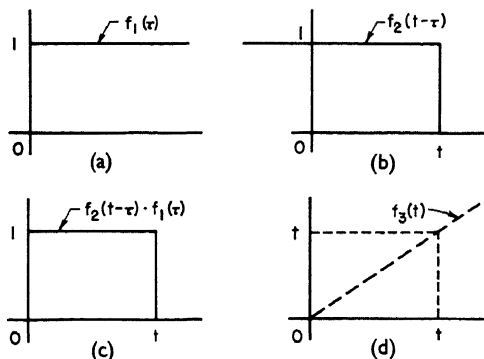


Fig. 10-4-2.

Example 10-4-2. The $f_1(t)$ and $f_2(t)$ functions for this example are shown in Fig. 10-4-3a and b respectively. If $f_1(t)$ is folded, the area under the product curve increases until $t = 1$. At this time $f_1(t)$ lies directly over $f_2(t)$. As $f_1(t)$ moves past

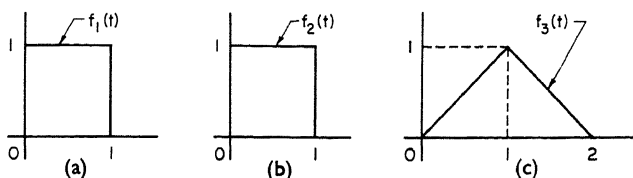


Fig. 10-4-3.

$t = 1$, the area under the product curve decreases linearly to a value of zero at $t = 2$. The resulting $f_3(t)$ function is shown in Fig. 10-4-3c.

Example 10-4-3. The previous example can be modified slightly to demonstrate an interesting point. In this example, both $f_1(t)$ and $f_2(t)$ are rectangular pulses

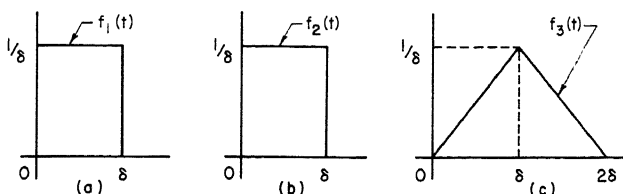


Fig. 10-4-4.

having a height of $1/\delta$ and a width of δ as shown in Fig. 10-4-4a and b respectively. The convolution of $f_1(t)$ and $f_2(t)$ is $f_3(t)$ as shown in Fig. 10-4-4c. In the limit as $\delta \rightarrow 0$, $f_1(t)$, $f_2(t)$, and $f_3(t)$ all lead to the unit impulse. Therefore it can be seen that the convolution of two unit impulses is a unit impulse.

When Laplace transforms are used, this result can be easily verified by the following steps

$$F_1(s) = \mathcal{L}[u_0(t)] = 1 \quad (10-4-3)$$

$$F_2(s) = \mathcal{L}[u_0(t)] = 1 \quad (10-4-4)$$

The product of the two transforms is

$$F_3(s) = F_1(s)F_2(s) = 1 \quad (10-4-5)$$

The $f_3(t)$ function is found as

$$f_3(t) = \mathcal{L}^{-1}[F_3(s)] = \mathcal{L}^{-1}[1] = u_0(t) \quad (10-4-6)$$

Example 10-4-4. The convolution of a unit impulse with some $f(t)$ results in the same $f(t)$. In explanation, $f_1(t)$ is taken as the unit impulse and $f_2(t)$ is any time function as shown in Fig. 10-4-5a and b respectively. When $f_1(t)$ is folded over, the same unit impulse results. As the impulse slides across $f_2(t)$, the product function is zero everywhere except at the point where the impulse is located. At this point the product function is an impulse which has a value equal to $f_2(t)$. Therefore the area under the product curve is equal to $f_2(t)$ for all values of t .

When Laplace transforms are used, this result can be easily verified by the following steps.

$$F_1(s) = \mathcal{L}[u_0(t)] = 1 \quad (10-4-7)$$

$$F_2(s) = \mathcal{L}[f_2(t)] \quad (10-4-8)$$

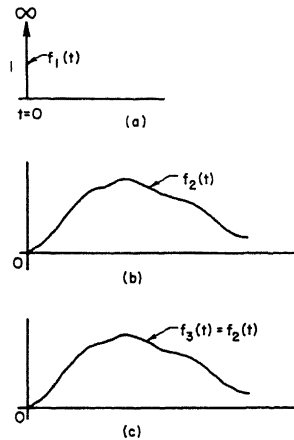


Fig. 10-4-5.

The product of the two transforms is

$$F_3(s) = F_1(s)F_2(s) = F_2(s) \quad (10-4-9)$$

The $f_3(t)$ function is found as

$$f_3(t) = \mathcal{L}^{-1}[F_3(s)] = \mathcal{L}^{-1}[F_2(s)] = f_2(t) \quad (10-4-10)$$

Example 10-4-5. The convolution of a unit doublet with some $f(t)$ results in the derivative of $f(t)$ as is demonstrated by depending on Laplace transforms in the following steps

$$F_1(s) = \mathcal{L}[u_{+1}(t)] = s \quad (10-4-11)$$

$$F_2(s) = \mathcal{L}[f_2(t)] \quad (10-4-12)$$

The product of the two transforms is

$$F_3(s) = F_1(s)F_2(s) = sF_2(s) \quad (10-4-13)$$

The $f_3(t)$ function is found as

$$f_3(t) = \mathcal{L}^{-1}[F_3(s)] = \mathcal{L}^{-1}[sF_2(s)] = \frac{d}{dt}[f_2(t)] \quad (10-4-14)$$

In this case, differentiation must be performed in the generalized sense.

As a generalization, the convolution of $u_{+n}(t)$ with some $f(t)$ results in the n th time derivative of the $f(t)$. Again each of the differentiations must be performed in the generalized sense.

Similarly, the convolution of $u_{-n}(t)$ with some $f(t)$ results in the n th integral of the $f(t)$.

10-5. Uses of the Convolution Integral for Physical Systems.

Fig. 10-5-1 is the block diagram for any physical system having the transfer function $T(s)$. The equation for $C(s)$ is

$$C(s) = R(s)T(s) \quad (10-5-1)$$

The time function $c(t)$ can be found as the inverse transform of $C(s)$ or as the convolution of $r(t)$ and $t(t)$

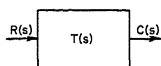


Fig. 10-5-1.

$$c(t) = \int_0^t r(t - \tau)t(\tau)d\tau \quad (10-5-2)$$

or

$$c(t) = \int_0^t r(\tau)t(t - \tau)d\tau \quad (10-5-3)$$

where $t(t)$ is the impulse response of the system. If the impulse response is known, the response of the system to any $r(t)$ input can be found by use of the convolution integral.

Eq. 10-5-1 can be modified as

$$C(s) = [sR(s)]\left[\frac{T(s)}{s}\right] \quad (10-5-4)$$

and the forms of the convolutions integral corresponding to this equation are

$$c(t) = \int_0^t r^1(t - \tau)t^{-1}(\tau)d\tau \quad (10-5-5)$$

or

$$c(t) = \int_0^t r^1(\tau)t^{-1}(t - \tau)d\tau \quad (10-5-6)$$

If the step function response is known, the response of the system to any $r(t)$ can be found by use of this form of the convolution integral.

Eq. 10-5-1 can also modified as

$$C(s) = [s^n R(s)]\left[\frac{T(s)}{s^n}\right] \quad (10-5-7)$$

and the corresponding integrals are

$$c(t) = \int_0^t r^n(t - \tau)t^{-n}(\tau)d\tau \quad (10-5-8)$$

or

$$c(t) = \int_0^t r^n(\tau)t^{-n}(t - \tau)d\tau \quad (10-5-9)$$

If the response of the system to an input $u_n(t)$ is known, the response of the system to any $r(t)$ can be found by use of these integrals.

In each set of the convolution integrals, the role of $r(t)$ and $t(t)$ can be interchanged.

Example 10-5-1. The following example is used to demonstrate these procedures. Fig. 10-5-2a is the block diagram for the system and Fig. 10-5-2b shows it in terms of the corresponding time functions.

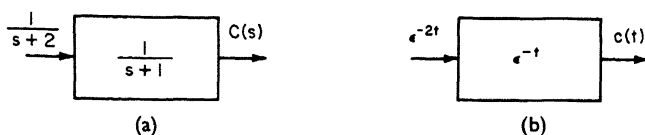


Fig. 10-5-2.

To check the results, the example is first worked by Laplace transforms. The transform of the output is

$$C(s) = R(s) T(s) = \frac{1}{(s+2)(s+1)} \quad (10-5-10)$$

and the inverse transform can be found as

$$c(t) = \mathcal{L}^{-1} \left[\frac{1}{(s+2)(s+1)} \right] = -\varepsilon^{-2t} + \varepsilon^{-t} \quad (10-5-11)$$

As an arbitrary choice, the convolution integral in the following form is used

$$c(t) = \int_0^t r^2(t-\tau) t^{-2}(\tau) d\tau \quad (10-5-12)$$

The first integral of $t(t)$ is

$$t^{-1}(t) = \int_0^t \varepsilon^{-t} dt = 1 - \varepsilon^{-t} \quad (10-5-13)$$

and the second integral is

$$t^{-2}(t) = \int_0^t (1 - \varepsilon^{-t}) dt = t - 1 + \varepsilon^{-t} \quad (10-5-14)$$

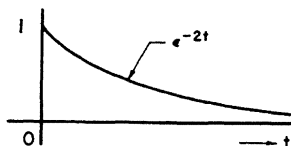


Fig. 10-5-3.

The input function $r(t)$ is sketched in Fig. 10-5-3 to visualize its derivative. At $t = 0$, $r(t)$ contains a unit discontinuity. Therefore $r^1(t)$ contains a unit impulse at $t = 0$, in addition to the conventional derivative.

$$r^1(t) = u_0(t) - 2\varepsilon^{-2t} \quad (10-5-15)$$

The second derivative is found as

$$r^2(t) = u_{+1}(t) - 2u_0(t) + 4\varepsilon^{-2t} \quad (10-5-16)$$

The concept of superposition is used and each of the terms of Eq. 10-5-16 is convolved with Eq. 10-5-14 separately and the resulting components of the solution are added. The notation used is that $c_1(t)$ results from $u_{+1}(t)$, $c_2(t)$ results from $-2u_0(t)$ and $c_3(t)$ results from $4\epsilon^{-2t}$. The total response is therefore

$$c(t) = c_1(t) + c_2(t) + c_3(t) \quad (10-5-17)$$

The component $c_1(t)$ results from convolving $u_{+1}(t)$ with $t^{-2}(t)$ and is therefore the derivative of $t^{-2}(t)$

$$c_1(t) = 1 - \epsilon^{-t} \quad (10-5-18)$$

The convolution of $-2u_0(t)$ with $t^{-2}(t)$ simply multiplies $t^{-2}(t)$ by -2

$$c_2(t) = -2t + 2 - 2\epsilon^{-t} \quad (10-5-19)$$

The component $c_3(t)$ results from convolving $4\epsilon^{-2t}$ with $t - 1 + \epsilon^{-t}$ as

$$\begin{aligned} c_3(t) &= \int_0^t 4\epsilon^{-2(t-\tau)} (\tau - 1 + \epsilon^{-\tau}) d\tau \\ &= -\epsilon^{-2t} + 2t - 3 + 4\epsilon^{-t} \end{aligned} \quad (10-5-20)$$

The total response is

$$c(t) = c_1(t) + c_2(t) + c_3(t) = -\epsilon^{-2t} + \epsilon^{-t} \quad (10-5-21)$$

It is not proposed that this is the best method of working the problem. The form of the convolution integral as given by Eq. 10-5-12 is used to demonstrate the manner in which the impulse family must be used to interpret differentiation properly.

10-6. Discussion of Fourier Transforms and the Convolution Integral. The existence of the Fourier transforms gives the designer a certain amount of assurance in that he knows the frequency response and the transient response are uniquely related. However, from the standpoint of design, the outlook is still not entirely satisfactory. The situation is again reviewed in which a system is to have a certain transient response but the designer accomplishes the design by using frequency response methods. As soon as the design is finished it is theoretically possible to go from the frequency response to the impulse response by use of the inverse Fourier transform and from the impulse response to the response to any other driving function by use of the convolution integral. The main problem involved in this approach is that if the resulting transient response is unsatisfactory, the designer has no clear-cut procedure for improving the design. About all that can be done is a cut-and-try method of redesigning the system. To put it mildly, this design method involves a great deal of computational effort.

Even if the problem faced by the designer is building a system to meet frequency response specification, the design procedure still has its uncertainties. For example, if the system is to be compensated, there is an infinite number of network that can be used. It can be seen intuitively that a certain compensation network or a family of networks may give

the best response. This design procedure is tolerable only because the designer does not attempt to find the best solution, he simply finds one solution satisfying the specifications. This implies that the system in all probability is potentially capable of satisfying more rigorous specifications than those demanded of it. In view of the high speed performance and accuracy required by industrial and military applications, it is regrettable if a system falls short of its full capabilities.

As a result of these difficulties other design methods have been developed. Their main objective is to execute the design so that both transient and sinusoidal frequency response are controlled simultaneously. In essence these design methods concentrate on the location of the poles and zeros of the closed-loop system. The manner in which the poles and zeros determine the transient and frequency response is introduced in a qualitative manner in Chapter 6 and is discussed in more quantitative terms in this chapter and in Chapter 13.

10-7. Transient and Frequency Response of Simple Systems.

In many cases, convolution can be used to investigate the response of systems with certain pole-zero patterns. The pole-zero pattern is divided into two portions such that the inverse transform of each portion is readily determined and the response of the original system is visualized as the convolution of these two time functions.

Before explaining the procedure, the time response of a simple pole-zero pattern

$$\frac{p(s+z)}{z(s+p)} \quad (10-7-1)$$

is explored because it is a useful tool in the discussion that follows.

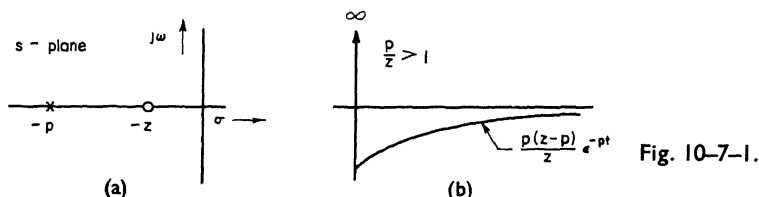
This pole-zero pattern can best be examined by expanding it as

$$\frac{p(s+z)}{z(s+p)} = \frac{p}{z} + \frac{p}{z} \frac{(z-p)}{(s+p)} \quad (10-7-2)$$

The corresponding time function is

$$\frac{p}{z} u_0(t) + \frac{p(z-p)}{z} \epsilon^{-pt} \quad (10-7-3)$$

This time function is considered in two separate cases.



For the first case, the zero is inside the pole as shown in Fig. 10-7-1a and the time response is as shown in Fig. 10-7-1b. Since p is larger than

z , the impulse has a value greater than unity and the coefficient of the exponential term is negative. The area under the entire curve is equal to unity.

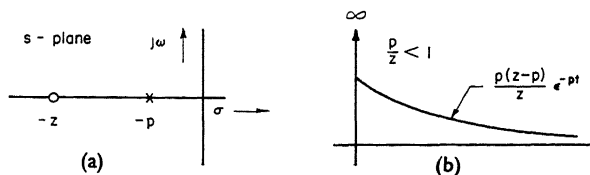


Fig. 10-7-2.

For the second case, the pole is inside the zero as shown in Fig. 10-7-2a and the time response is as shown in Fig. 10-7-2b. Since p is smaller than z , the impulse has a value less than unity and the coefficient of the exponential term is positive. The area under the entire curve is equal to unity.

Example 10-7-1. The pole-zero configuration to be investigated is

$$T(s) = \frac{C(s)}{R(s)} = \frac{3(s+2)}{2(s+1)(s+3)} \quad (10-7-4)$$

The response to a unit-step function is obtained from

$$C(s) = \frac{3(s+2)}{2s(s+1)(s+3)} \quad (10-7-5)$$

This $C(s)$ is divided into two portions in two different ways. First, it is divided into $C_1(s)$ and $C_2(s)$ as

$$C_1(s) = \frac{1}{s(s+1)} \quad (10-7-6)$$

and

$$C_2(s) = \frac{3(s+2)}{2(s+3)} \quad (10-7-7)$$

Second, it is divided into $C_3(s)$ and $C_4(s)$ as

$$C_3(s) = \frac{3}{s(s+3)} \quad (10-7-8)$$

and

$$C_4(s) = \frac{(s+2)}{2(s+1)} \quad (10-7-9)$$

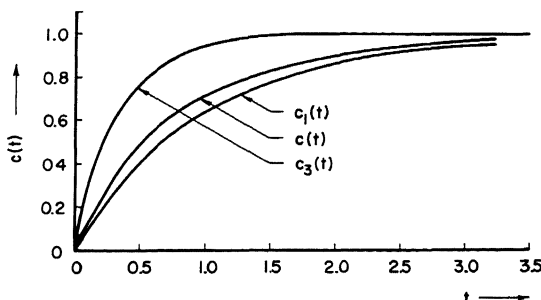


Fig. 10-7-3.

The time functions $c_1(t)$ and $c_3(t)$ are readily found as

$$c_1(t) = 1 - e^{-t} \quad (10-7-10)$$

and

$$c_3(t) = 1 - e^{-3t} \quad (10-7-11)$$

these curves are shown in Fig. 10-7-3.

The function $c_1(t)$ is convolved with $c_2(t)$ to obtain $c(t)$. $C_2(s)$ has its zero inside the pole; therefore, $c_2(t)$ contains an impulse of value greater than unity and a negative exponential term as

$$c_2(t) = \frac{3}{2}u_0(t) - \frac{3}{2}e^{-3t} \quad (10-7-12)$$

The convolution of $c_1(t)$ and $c_2(t)$ can be visualized by folding $c_2(t)$ and letting it slide across $c_1(t)$. The impulse has the effect of multiplying $c_1(t)$ by $3/2$, but the negative exponential term holds $c(t)$ to a value of unity at $t = \infty$. Therefore, it can be seen that $c(t)$ is above $c_1(t)$.

The function $c_3(t)$ is convolved with $c_4(t)$ to obtain $c(t)$. $C_4(s)$ has its pole inside the zero, therefore $c_4(t)$ contains an impulse of value less than unity and a positive exponential term as

$$c_4(t) = \frac{1}{2}u_0(t) + \frac{1}{2}e^{-t} \quad (10-7-13)$$

The convolution of $c_3(t)$ and $c_4(t)$ can be visualized by folding $c_4(t)$ and letting it slide across $c_3(t)$. The impulse has the effect of multiplying $c_3(t)$ by $1/2$, but the positive exponential term brings $c(t)$ to a value of unity at $t = \infty$. Therefore it can be seen that $c(t)$ is below $c_3(t)$.

When the preceding two discussions are considered simultaneously, it can be seen that $c(t)$ lies between $c_1(t)$ and $c_3(t)$. The $c(t)$ curve on Fig. 10-7-3 is computed to verify this result.

Example 10-7-2. This example uses the closed-loop transfer function of Ex. 10-7-1 as given by

$$T(s) = \frac{C(s)}{R(s)} = \frac{3(s+2)}{2(s+1)(s+3)} \quad (10-7-14)$$

However, in this example the sinusoidal steady-state response is to be investigated

$$T(j\omega) = \frac{C(j\omega)}{R(j\omega)} = \frac{3(j\omega+2)}{2(j\omega+1)(j\omega+3)} \quad (10-7-15)$$

In order to analyze the systems in a manner analogous to the previous example, the simple pole-zero pattern

$$\frac{p(j\omega+z)}{z(j\omega+p)} \quad (10-7-16)$$

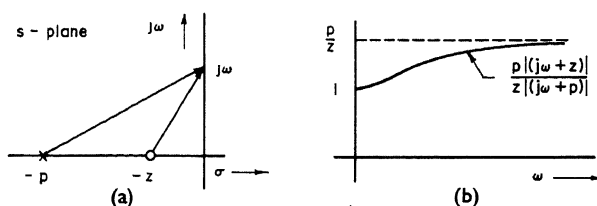


Fig. 10-7-4.

is first explored. When the zero is inside the pole, the frequency response starts at unity and increases to the value p/z as shown in Fig. 10-7-4b. When the pole is inside the zero, the frequency response starts at unity and decreases to the value

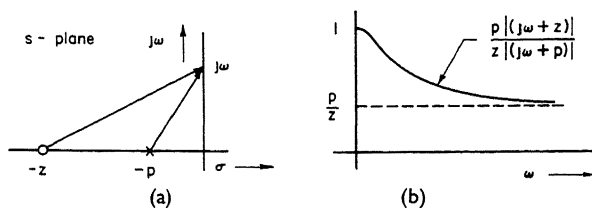


Fig. 10-7-5.

p/z as shown in Fig. 10-7-5b. A study of the lengths of the phasors in Figs. 10-7-4a and 10-7-5a indicates why the two curves have their respective shapes.

The $T(j\omega)$ of Eq. 10-7-15 is divided into two portions in two different ways. First it is divided into $T_1(j\omega)$ and $T_2(j\omega)$ as

$$T_1(j\omega) = \frac{1}{j\omega + 1} \quad (10-7-17)$$

and

$$T_2(j\omega) = \frac{3(j\omega + 2)}{2(j\omega + 3)} \quad (10-7-18)$$

Second it is divided into $T_3(j\omega)$ and $T_4(j\omega)$ as

$$T_3(j\omega) = \frac{3}{j\omega + 3} \quad (10-7-19)$$

and

$$T_4(j\omega) = \frac{j\omega + 2}{2(j\omega + 1)} \quad (10-7-20)$$

The functions $|T_1(j\omega)|$ and $|T_3(j\omega)|$ are shown in Fig. 10-7-6.

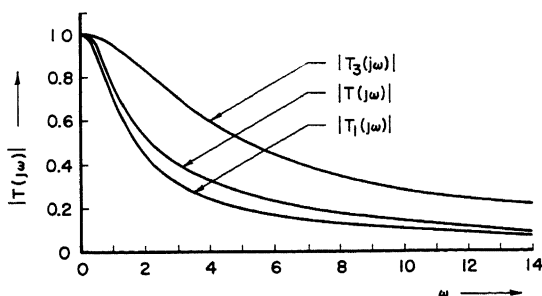


Fig. 10-7-6.

When $|T_1(j\omega)|$ is multiplied by $|T_2(j\omega)|$, the curve $|T(j\omega)|$ results. It can be seen that $|T(j\omega)|$ is above $|T_1(j\omega)|$ because the zero of $F_2(s)$ is inside the pole. When $|T_3(j\omega)|$ is multiplied by $|T_4(j\omega)|$, the curve $|T(j\omega)|$ again results. It can be seen that $|T(j\omega)|$ is below $|T_3(j\omega)|$ because the pole of $F_4(s)$ is inside the zero.

When the preceding two discussions are considered simultaneously, it is seen that the $|T(j\omega)|$ lies between $|T_1(j\omega)|$ and $|T_3(j\omega)|$. The $|T(j\omega)|$ curve of Fig. 10-7-6 is computed to verify this result.

Example 10-7-3. This example is similar to the preceding two except that the zero is a variable

$$T(s) = \frac{C(s)}{R(s)} = \frac{3(s+z)}{s(s+1)(s+3)} \quad (10-7-21)$$

The transform for the unit-step response is

$$C(s) = \frac{3(s+z)}{s(s+1)(s+3)} \quad (10-7-22)$$

If the zero is between the poles at -1 and -3 , the $c(t)$ response is bracketed between $c_1(t)$ and $c_3(t)$ of Fig. 10-7-3. If the zero is at -3 , the zero cancels the pole at -3 and the $c(t)$ is equal to $c_1(t)$. As the zero moves toward -1 , $c(t)$ approaches $c_3(t)$ and exactly equals $c_3(t)$ when the zero is at -1 .

The frequency response for the closed-loop system is given by

$$T(j\omega) = \frac{3(j\omega+z)}{z(j\omega+1)(j\omega+3)} \quad (10-7-23)$$

If the zero is between the poles at -1 and -3 , the $|T(j\omega)|$ curve is bracketed between $|T_1(j\omega)|$ and $|T_3(j\omega)|$ of Fig. 10-7-6. If the zero is at -3 , the zero cancels the pole at -3 and $|T(j\omega)|$ is equal to $|T_1(j\omega)|$. As the zero moves toward -1 , $|T(j\omega)|$ approaches $|T_3(j\omega)|$ and exactly equals $|T_3(j\omega)|$ when the zero is at -1 .

The zero is carried inside the pole at -1 in the next example.

Example 10-7-4. This example resembles the preceding three except that the zero is inside the pole at -1 .

$$T(s) = \frac{C(s)}{R(s)} = \frac{6(s+0.5)}{(s+1)(s+3)} \quad (10-7-24)$$

The transform for the unit-step response is

$$C(s) = \frac{6(s+0.5)}{s(s+1)(s+3)} \quad (10-7-25)$$

The corresponding time function is

$$c(t) = 1 + \frac{3}{2}e^{-t} - \frac{5}{2}e^{-3t} \quad (10-7-26)$$

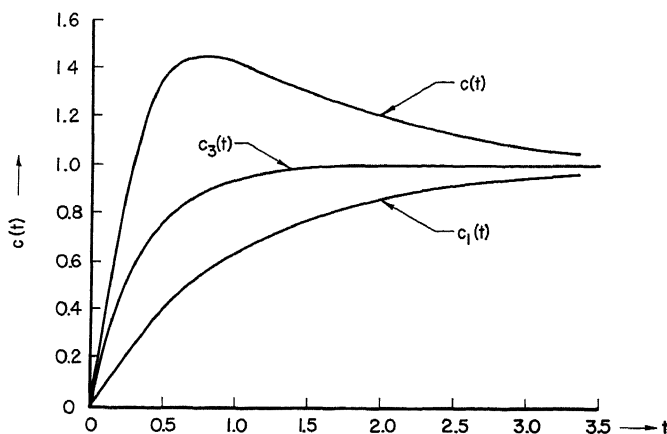


Fig. 10-7-7.

The time function $c(t)$ is plotted in Fig. 10-7-7 along with $c_1(t)$ and $c_3(t)$ from Fig. 10-7-3 which are added for the sake of reference.

The frequency response of the closed-loop system is given by

$$T(j\omega) = \frac{6(j\omega + 0.5)}{(j\omega + 1)(j\omega + 3)} \quad (10-7-27)$$

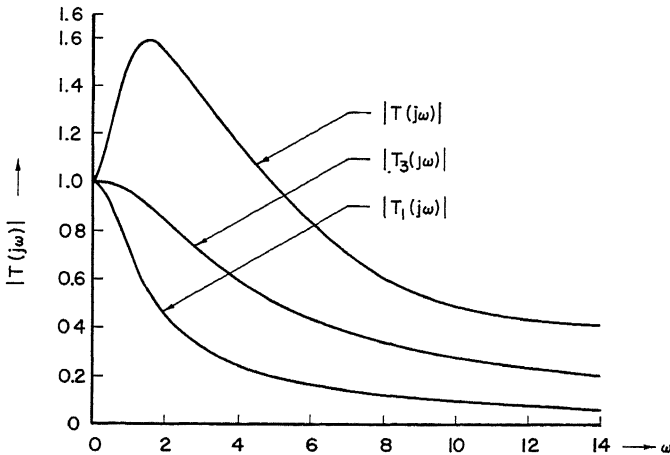


Fig. 10-7-8.

The resulting $|T(j\omega)|$ curve is shown in Fig. 10-7-8 along with $|T_1(j\omega)|$ and $|T_3(j\omega)|$ from Fig. 10-7-6 which are added for the sake of reference.

It is interesting to note that the step function response has an overshoot of approximately 45% and the frequency response has a maximum value of approximately 1.58 even though the system's poles and zeros are on the negative real axis. Here is an example in which the system's frequency response is peaked and yet the system's transient response is not oscillatory; however, the transient response does overshoot its steady-state value.

Example 10-7-5. Some of the concepts discussed in the previous examples are used to analyze the transform of the unit-step response as given by

$$C(s) = \frac{1.83(s + 1.5)(s + 2.5)(s + 3.5)}{s(s + 1)(s + 2)(s + 3)(s + 4)} \quad (10-7-28)$$

This pole-zero pattern is divided into two portions in two different ways. First it is divided into

$$C_1(s) = \frac{1}{s(s + 1)} \quad (10-7-29)$$

$C_2(s)$ being composed of the remaining poles and zeros. Second it is divided into

$$C_3(s) = \frac{4}{s(s + 4)} \quad (10-7-30)$$

$C_4(s)$ being composed of the remaining poles and zeros.

The function $c_1(t)$ is

$$c_1(t) = 1 - e^{-t} \quad (10-7-31)$$

The poles and zeros of $C_2(s)$ can be divided into three pole-zero pairs as shown in Fig. 10-7-9a. Each pair has the zero inside the pole and $c_1(t)$ can be convolved with each pair separately. Therefore, $c(t)$ is above $c_1(t)$.

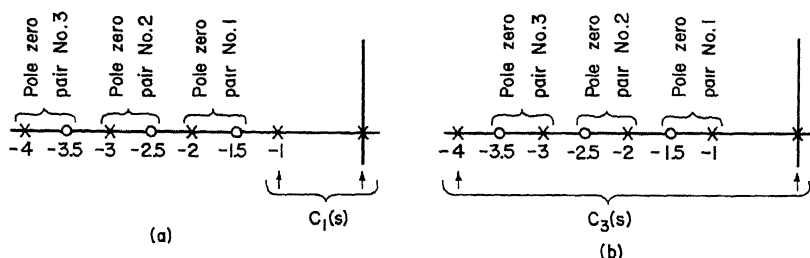


Fig. 10-7-9.

The function $c_3(t)$ is

$$c_3(t) = 1 - e^{-4t} \quad (10-7-32)$$

The poles and zeros of $C_4(s)$ can be divided into three pole-zero pairs as shown in Fig. 10-7-9b. Each pair has the pole inside the zero and $c_3(t)$ can be convolved with each pair separately.

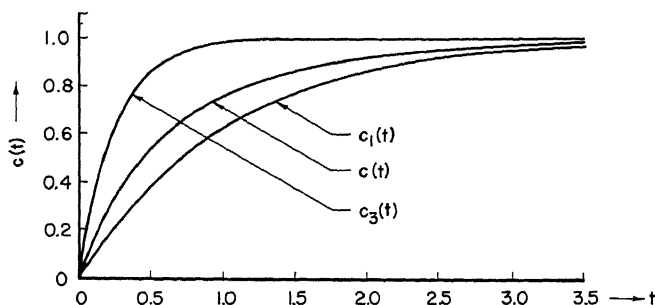


Fig. 10-7-10.

It can be seen that $c(t)$ is between $c_3(t)$ and $c_1(t)$. The $c(t)$ curve of Fig. 10-7-10 is computed to verify this result.

Example 10-7-6. It should not be inferred from the preceding examples that a zero must be closer to the origin than any pole in order for the step function response to overshoot its steady-state value. For example, the transform of a unit-step response is given by

$$C(s) = \frac{24.8(s + 1.1)^2}{s(s + 1)(s + 3)(s + 10)} \quad (10-7-33)$$

This transform has a second order zero at -1.1 . It can be reasoned that one of these zeros essentially cancels the pole at -1 and the system behaves as if it has a simple zero inside the pole at -3 . The time function $c(t)$ is shown in Fig. 10-7-11.

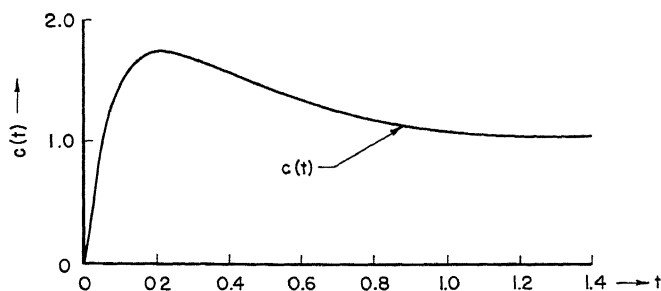


Fig. 10-7-11.

The frequency response of the system is obtained from

$$T(j\omega) = \frac{24.8(j\omega + 1.1)^2}{(j\omega + 1)(j\omega + 3)(j\omega + 10)} \quad (10-7-34)$$

and the plot of this function is shown in Fig. 10-7-12.

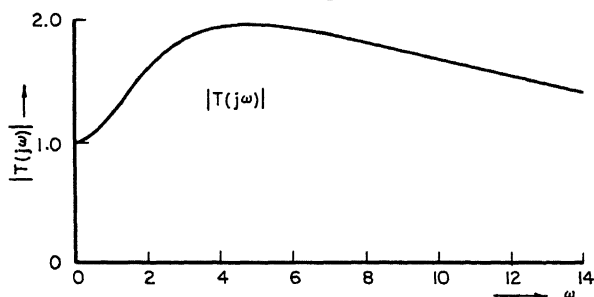


Fig. 10-7-12.

Example 10-7-7. This example starts with the step function transform $C(s) = \frac{1}{s(s^2 + s + 1)}$ and adds a pole-zero pair in two ways as

$$C_1(s) = \frac{2(s+1)}{(s+2)} C(s) = \frac{2(s+1)}{s(s+2)(s^2 + s + 1)} \quad (10-7-35)$$

and

$$C_2(s) = \frac{0.5(s+2)}{(s+1)} C(s) = \frac{0.5(s+2)}{s(s+1)(s^2 + s + 1)} \quad (10-7-36)$$

The time function $c(t)$ can be obtained from the generalized curves of Chapter 6 with $\zeta = 0.5$ and by letting $\omega_n = 1$. This $c(t)$ is shown in Fig. 10-7-13. The $c(t)$ is convolved with a time function having a transform with a zero inside the pole. Therefore, this time function contains an impulse of value greater than unity and a negative exponential term as

$$2u_0(t) - 2e^{-2t} \quad (10-7-37)$$

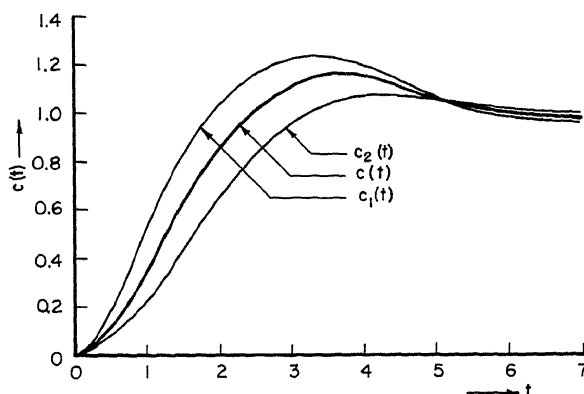


Fig. 10-7-13.

The impulse has the effect of multiplying $c(t)$ by 2 but the negative exponential term reduces $c_1(t)$ to a value of unity at $t = \infty$. The function $c_1(t)$ is shown in Fig. 10-7-13. To obtain $c_2(t)$, $c(t)$ is convolved with a time function which has a transform with the pole inside the zero. Therefore, this time function contains an impulse of value less than unity and a positive exponential term as

$$\frac{1}{2}u_0(t) + \frac{1}{2}e^{-t} \quad (10-7-38)$$

The impulse has the effect of multiplying $c(t)$ by 1/2 but the positive exponential term brings $c_2(t)$ to a value of unity at $t = \infty$. The function $c_2(t)$ is also shown in Fig. 10-7-13.

The frequency responses of these three systems are obtained from

$$T(j\omega) = \frac{1}{1 - \omega^2 + j\omega} \quad (10-7-39)$$

$$T_1(j\omega) = \frac{2(j\omega + 1)}{(j\omega + 2)[1 - \omega^2 + j\omega]} \quad (10-7-40)$$

and

$$T_2(j\omega) = \frac{0.5(j\omega + 2)}{(j\omega + 1)[1 - \omega^2 + j\omega]} \quad (10-7-41)$$

and the plots of these functions are shown in Fig. 10-7-14.

A review of this example indicates that adding a pole-zero pair with the zero inside the pole tends to make the step function response build up more rapidly, increases the overshoot, and builds up the frequency response to a higher peak with an increased bandwidth. When the pole-zero pair has the pole inside the zero, the corresponding step function response builds up more slowly, has a decreased overshoot, and the frequency response builds up to a lower peak with a smaller bandwidth.

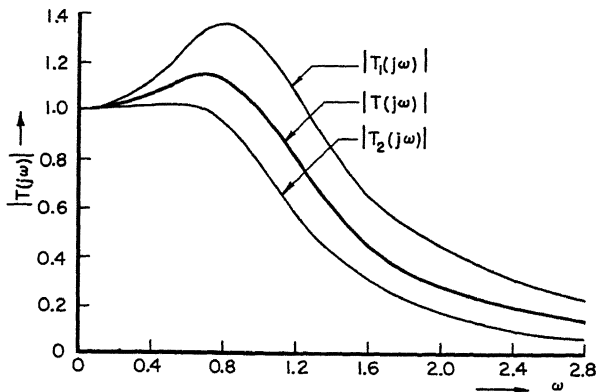


Fig. 10-7-14.

Example 10-7-8. The effects of the movement of the pole and zero on the step function response are shown more completely in the six parts of Fig. 10-7-15. In each of the first five parts, the pole remains fixed and the zero is moved as indicated on the curves. In each case, the closer the zero is to the origin the greater is the overshoot. When the zero cancels the pole, the transform

$$C_0(s) = \frac{1}{s(s^2 + s + 1)} \quad (10-7-42)$$

results and $c_0(t)$, the corresponding inverse transform is shown with a heavy line on each part of the figure.

The curves of Fig. 10-7-15f are for the zero fixed at -1 and pole allowed to move as indicated. The closer the pole to the origin, the less the overshoot. The curves of Fig. 10-7-15f are obtained by replotting selected curves from the first five parts of Fig. 10-7-15. Other sets of curves can also be selected to show the effects of the movement of the pole for other zero locations but these are left as Probs. 10-30 through 10-33.

All the curves in Fig. 10-7-15 are variations of the system described by

$$C_0(s) = \frac{1}{s(s^2 + s + 1)} \quad (10-7-43)$$

as affected by the addition of a pole-zero pair. This original system is the $\zeta = 0.5$ and $\omega_n = 1$ curve from Fig. 6-11-2. By the use of scaling, ω_n can be changed from 1 to any other desired value. However, for ζ to be changed an entire new set of curves must be calculated. The relative effects of the movements of poles and zeros remain the same and the curves of Fig. 10-7-15 can be used as a guide in estimating the new system response. Actually, as discussed in Chapter 13, this system offers the designer a great deal of flexibility and systems based on these curves satisfy a wide variety of specifications.

10-8. Maximally Flat Functions. If the curves of Fig. 6-23-1 are inspected, it is seen that the frequency response curve for $\zeta = 0.707$ has

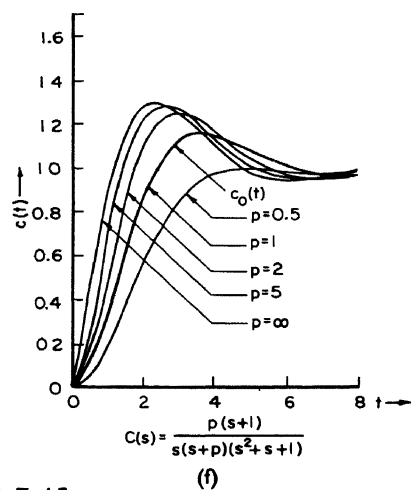
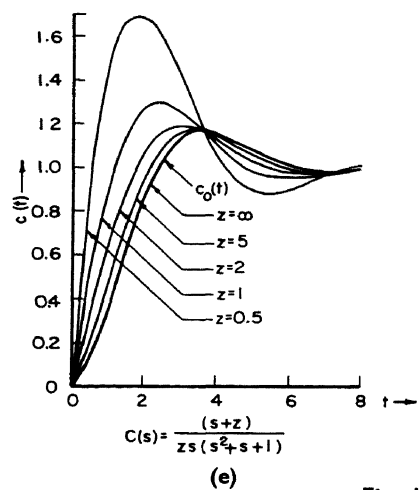
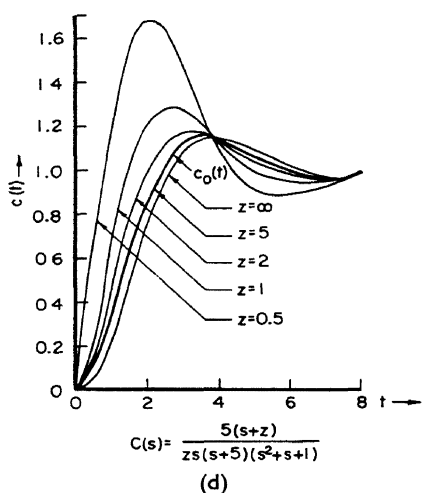
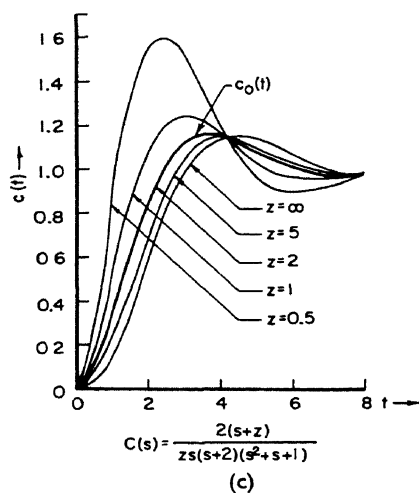
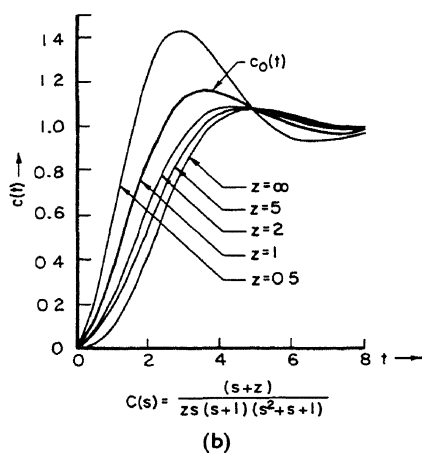
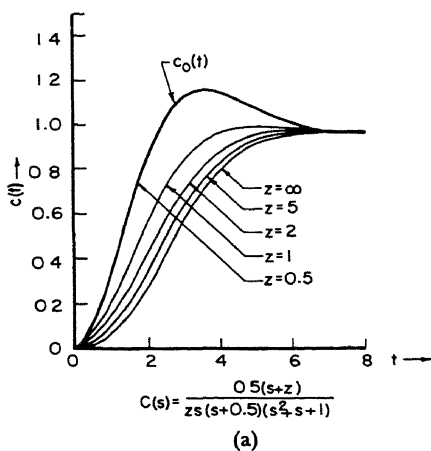


Fig. 10-7-15.

zero slope at $\omega = 0$. This frequency function is one member of a family of functions called maximally flat functions or Butterworth functions.

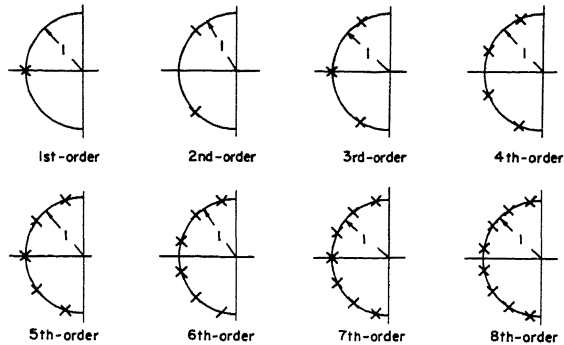


Fig. 10-8-1.

The pole locations of these functions are shown in Fig. 10-8-1 through the eighth order. From this figure, it can be seen that the system with $\zeta = 0.707$ of Fig. 6-23-1 is a second-order function. For the first four such functions, the closed-loop transfer functions are given by

$$\text{1st-order } T_1(s) = \frac{1}{(s + 1)} \quad (10-8-1)$$

$$\text{2nd-order } T_2(s) = \frac{1}{(s^2 + \sqrt{2}s + 1)} \quad (10-8-2)$$

$$\text{3rd-order } T_3(s) = \frac{1}{(s + 1)(s^2 + s + 1)} \quad (10-8-3)$$

$$\text{4th-order } T_4(s) = \frac{1}{(s^2 + 1.8478s + 1)(s^2 + 0.7654s + 1)} \quad (10-8-4)$$

If $j\omega$ is substituted for s , the magnitude squared of the n th-order function can be shown to be

$$|T_n(j\omega)|^2 = \frac{1}{1 + \omega^{2n}} \quad (10-8-5)$$

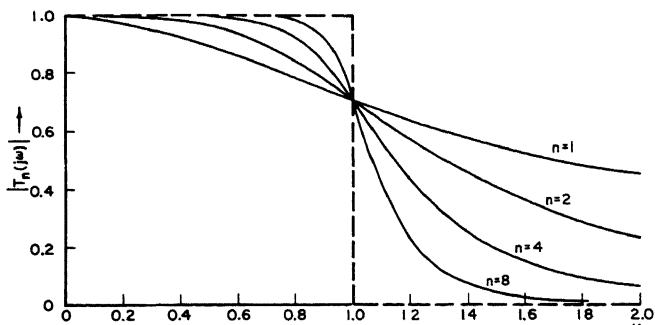
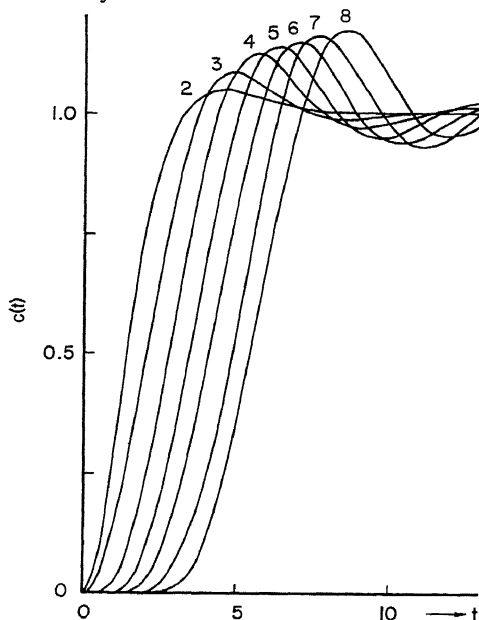


Fig. 10-8-2.

A few of these functions are shown in Fig. 10-8-2. The function being approximated as n increases is the idealized low-pass filter characteristic

shown by dotted lines. It can be shown that for a given n , the first $2n-1$ derivatives of $|T_n(j\omega)|^2$ are zero at the origin and this is why the function is known as maximally flat.



(From D. Graham and R. C. Lathrup, "The Synthesis of 'Optimum' Transient Response," *Trans AIEE*, 72, Part II, *Applications and Industry* (1953) 273, with permission)

Fig. 10-8-3.

The step function responses of the second- through eighth-order Butterworth functions are shown in Fig. 10-8-3. It is interesting to note that this family of functions does not have peaked frequency response curves and yet the transient response is oscillatory. For the eighth-order function, the step function response has approximately a 19% overshoot. It is also interesting to note that if only the poles nearest the $j\omega$ -axis for the eighth-order system were present, the system would have a $\zeta = \cos 11.25^\circ = 0.195$. Based on Fig. 6-11-3 the step function response would have an overshoot of 55% and based on Fig. 6-23-2 the frequency response would have a value for M_m of 2.5. If the other six poles are thought of as being brought up to their locations as a group the net effect is to decrease the overshoot to the 19% value and M_m to 1.0.

In order to understand how the poles of a given system can be complex and yet the frequency response curves not be peaked, the second-order function is studied. The function being considered is

$$T(j\omega) = \frac{1}{(1 - \omega^2) + j1.414\omega}$$

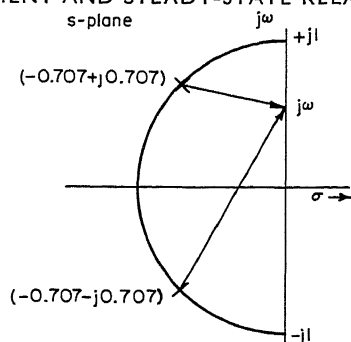


Fig. 10-8-4.

The pole location and the phasor for a specific value of $j\omega$ are shown in Fig. 10-8-4. As ω increases from zero, the phasor from the upper pole decreases in length. At the same time the phasor from the lower pole increases in length in such a manner that the product of the two lengths remains essentially constant for low frequencies and gradually increases. As $j\omega$ passes $+j0.707$, both phasors increase in length and the product of the two lengths begins to increase rapidly. Since this phasor product is in the denominator of $T(j\omega)$, the shape of the curves for the second-order system in Fig. 10-8-2 results. This same type of reasoning can be extended to functions of higher order.

The maximally flat functions may not be encountered very often in control systems; however, they are discussed here in order to increase the designer's over-all understanding of pole-zero patterns.

10-9. Non-minimum Phase Functions. For a function to represent a stable driving point impedance, the poles must be in the left half plane or on the $j\omega$ -axis. This is also true of the zeros because the function can be "flipped over," and it then represents a driving point admittance which must also be stable. For a control system to be classed as stable, the poles must all be to the left of the $j\omega$ -axis. For a transfer function to represent a stable system, all the poles must be in the left half plane; however, the zeros are not necessarily in the left half plane.

The possibility of having zeros in the right half plane brings up the interesting fact that a given attenuation function may have more than one possible pole-zero pattern. This point is best explained by starting with several pole-zero patterns that yield the same attenuation curves. Four such patterns are shown in Fig. 10-9-1. As can be seen, the straight line approximation for all patterns breaks up by 20 db/decade at $\omega = 1/2$, down by 60 db/decade at $\omega = 1$, and then up by 20 db/decade at $\omega = 2$.

The only difference among these four patterns is in the phase characteristics. The patterns of Fig. 10-9-1a and d are compared in the following discussion. In Fig. 10-9-1a, since all three poles are in the left half plane, the net contribution to the phase from the poles in going from $\omega = 0$ to $\omega = \infty$ is a rotation in the counterclockwise direction of $3 \times 90^\circ$

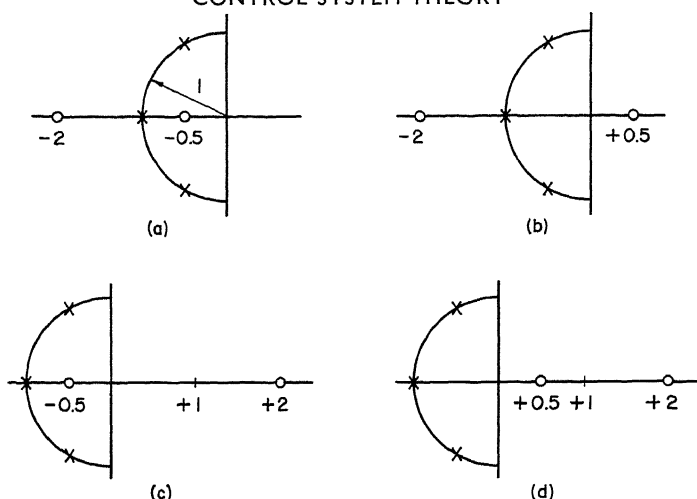


Fig. 10-9-1.

in the denominator of the transfer function. Since both zeros are in the left half plane, the net contribution to the phase from $\omega = 0$ to $\omega = \infty$ is a rotation in the counterclockwise direction of $2 \times 90^\circ$ in the numerator of the transfer function. When the angle is brought from the denominator to the numerator, the total change in phase is equal to 90° in the clockwise direction. In Fig. 10-9-1d, the net rotation from the poles remains the same but the rotation from the zeros is now in the clockwise direction and the total change in phase is equal to $5 \times 90^\circ$ in the clockwise direction.

When all the possible patterns are compared, it can be seen that a transfer function with all its zeros in the left half plane has a smaller change in phase as ω goes from zero to ∞ than any of the other patterns with the same magnitude characteristic and is referred to as minimum phase. If the transfer function has a zero (or zeros) in the right half plane it is referred to as non-minimum phase.

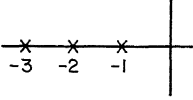
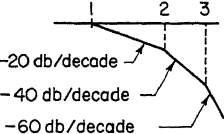
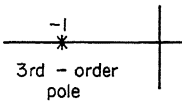
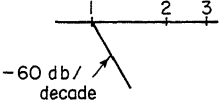
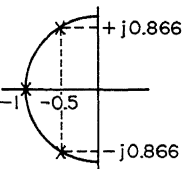
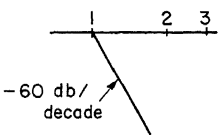
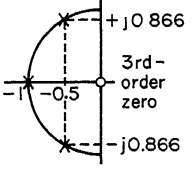
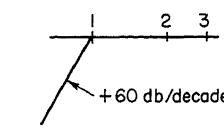
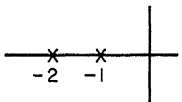
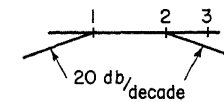
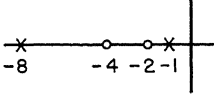
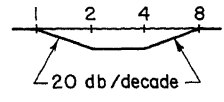
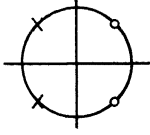
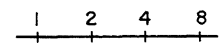
For a given transfer function to have a zero in the right half plane, the signal somewhere in the system must travel along parallel paths and recombine to create the zero. Examples of this are discussed in Chapter 12. The converse of this is not necessarily true. That is, a system can have parallel paths without necessarily having a zero in the right half plane.

If a function is known to be minimum phase, then for a given attenuation curve there is only one possible phase function and by use of the concept of the inverse Fourier transform there is only one possible impulse response function. When the convolution integral is added as a tool, the response to any other driving function can be uniquely determined as soon as the impulse response is found.

Although the transfer functions coming from control systems are not necessarily minimum phase, they usually are. For this reason most of the control system literature (as in this book) is concerned with minimum phase functions.

10-10. Other Pole-Zero Patterns. The usual control system acts as if it were a low-pass filter. Therefore most of the pole-zero patterns discussed in this article are not typical of control systems but are explored to add to the designer's over-all understanding of pole-zero patterns. The entire discussion refers to Table 10-10-1.

TABLE 10-10-1

Transfer function	Pole - zero pattern	Asymptotes for attenuation diagram	Characteristics
(a) $\frac{6}{(s+1)(s+2)(s+3)}$			Low pass filter
(b) $\frac{1}{(s+1)^3}$			Low pass filter
(c) $\frac{1}{(s+1)(s^2+s+1)}$			Low pass filter maximum flat
(d) $\frac{s^3}{(s+1)(s^2+s+1)}$			High pass filter
(e) $\frac{2s}{(s+1)(s+2)}$			Band pass filter
(f) $\frac{(s+2)(s+4)}{(s+1)(s+8)}$			Band elimination filter
(g) $\frac{s^2 - \sqrt{2}s + 1}{s^2 + \sqrt{2}s + 1}$			All pass filter

Function (a) in Table 10-10-1 is the transfer function of a low-pass filter with the asymptote for the attenuation diagram breaking down with a change in slope of 20 db/decade at $\omega = 1, 2$, and 3.

Transfer function (b) also describes a low-pass filter but the three poles have been moved together at -1 and the asymptotes for the attenuation diagram break down by 60 db/decade at $\omega = 1$.

Transfer function (c) is a third-order Butterworth function and the asymptotes for the attenuation diagram break down by 60 db/decade at $\omega = 1$. The asymptotes for functions (b) and (c) are identical but the actual attenuation curves differ. This is explored in Prob. 10-35.

Function (d) can be obtained from function (c) by the change in variable $s = 1/s$; therefore function (d) describes a high-pass filter.

Function (e) is typical (except for a gain constant) of a single-stage R-C coupled amplifier and the resulting response is that of a band-pass filter as shown.

Because of the arrangement of the poles and zeros of function (f), the asymptotes break down at $\omega = 1$, then up at $\omega = 2$, and $\omega = 4$ and down at $\omega = 8$. The result is the band-elimination characteristic as shown.

If a pole-zero pattern has a zero in the right half plane that is located as the mirror image with respect to the $j\omega$ -axis for each pole in the left half plane, then the resulting function describes an all-pass filter. Such a pole-zero pattern is given as function (g). At each ω that a pole would cause an asymptote to break down, the corresponding zero cancels this effect by an equal break up. All-pass filters, therefore, are non-minimum phase. Since the all-pass filter does not cause attenuation at any frequency its total effect is one of phase shift.

10-11. Error Series and Relations to the Closed-Loop Pole-Zero Location. In Chapter 6, the static error coefficients for a unity feedback system are introduced. The position constant K_p is defined as

$$K_p = \lim_{s \rightarrow 0} G(s) \quad (10-11-1)$$

the velocity constant K_v , as

$$K_v = \lim_{s \rightarrow 0} sG(s) \quad (10-11-2)$$

and the acceleration constant as

$$K_a = \lim_{s \rightarrow 0} s^2 G(s) \quad (10-11-3)$$

These constants are of considerable importance, since they are related to the steady-state errors in Type 0, 1, and 2 systems. In some cases more information about the system error than simply K_p , K_v , and K_a is needed. A so-called error series is available for a more extended study of system errors. This series is obtained by expanding $E(s)/R(s)$ in a power series in powers of s as

$$\frac{E(s)}{R(s)} = C_0 + C_1s + C_2s^2 + \cdots \quad (10-11-4)$$

Under certain conditions K_p , K_v , and K_a are related to C_0 , C_1 , and C_2 respectively. To illustrate this, Eq. 10-11-4 is studied first for a Type 0 unity feedback system.

$$\frac{E(s)}{R(s)} = \frac{1}{1 + G(s)} = C_0 + C_1s + C_2s^2 + \cdots \quad (10-11-5)$$

From Eq. 10-11-5 for a Type 0 system

$$C_0 = \lim_{s \rightarrow 0} \frac{E(s)}{R(s)} = \frac{1}{1 + \lim_{s \rightarrow 0} G(s)} = \frac{1}{1 + K_p} \quad (10-11-6)$$

For a type 1 system K_p is infinite and C_0 is seen to be zero. For a Type 1 unity feedback system, then,

$$\frac{E(s)}{R(s)} = C_1s + C_2s^2 + \cdots = \frac{1}{1 + G(s)} \quad (10-11-7)$$

$$C_1 = \lim_{s \rightarrow 0} \frac{E(s)}{sR(s)} = \lim_{s \rightarrow 0} \frac{1}{s + sG(s)} = \frac{1}{K_v} \quad (10-11-8)$$

If K_v is infinite, this procedure can be extended to show that

$$C_2 = \frac{1}{K_a} \quad (10-11-9)$$

The constants C_0 , C_1 , and C_2 can be found in general by any scheme that realizes the series form of Eq. 10-11-4.

Example 10-11-1. The error series is to be found for a general second-order, unity feedback, Type 1 system.

$$G(s) = \frac{K}{s(s + a)} \quad (10-11-10)$$

$$\frac{E(s)}{R(s)} = \frac{s^2 + as}{s^2 + as + K} = \frac{a}{K}s + \left(1 - \frac{a^2}{K}\right)s^2 + \cdots \quad (10-11-11)$$

The error series is easily found by long division as follows

$$K + as + s^2 \overline{\left| \begin{array}{l} (a/K)s + (1 - a^2/K)s^2 + \cdots \\ as + \frac{a^2}{K}s^2 + \frac{a}{K}s^3 \\ \hline \left(1 - \frac{a^2}{K}\right)s^2 - \frac{a}{K}s^3 \end{array} \right.} \quad (10-11-12)$$

An alternate scheme for finding C_0 , C_1 , etc., is based on the theory of the Maclaurin series from which

$$\begin{aligned} C_0 &= \frac{E(0)}{R(0)} \\ C_1 &= \frac{d}{ds} \left[\frac{E(s)}{R(s)} \right] \bigg|_{s=0} \\ C_2 &= \frac{1}{2!} \frac{d^2}{ds^2} \left[\frac{E(s)}{R(s)} \right] \bigg|_{s=0} \end{aligned} \quad (10-11-13)$$

One additional aspect of the error series is of use. This aspect is the relationship between the coefficients in the error series and the closed-loop poles and zeros. For a unity feedback system.

$$T(s) = \frac{C(s)}{R(s)} = 1 - \frac{E(s)}{R(s)} = 1 - C_0 - C_1 s - C_2 s^2 - \cdots \quad (10-11-14)$$

$$= \frac{K(s^m + a_{m-1}s^{m-1} + \cdots + a_1 s + a_0)}{s^n + b_{n-1}s^{n-1} + \cdots + b_1 s + b_0} \quad (10-11-15)$$

A long division process similar to that used in Ex. 10-11-1 gives

$$T(s) = K \left(\frac{a_0}{b_0} \right) + \frac{K}{b_0} \left(a_1 - \frac{b_1 a_0}{b_1} \right) s + \cdots \quad (10-11-16)$$

From Eqs. 10-11-14 and -16

$$1 - C_0 = K \left(\frac{a_0}{b_0} \right) = \frac{K_p'}{1 + K_p'} \quad (10-11-17)$$

$$-C_1 = \frac{K}{b_0} \left(a_1 - \frac{b_1 a_0}{b_1} \right) = -\frac{1}{K_v'} \quad (10-11-18)$$

in which K_p' and K_v' are dynamic error constants and reduce to the static error constant only in case the one of next lower order is infinite. The correlation of K_p' and K_v' with closed-loop poles and zeros is obtained by writing

$$T(s) = \frac{K(s + z_1)(s + z_2) \cdots (s + z_m)}{(s + p_1)(s + p_2) \cdots (s + p_n)} \quad (10-11-19)$$

and noting that a_0 in Eq. 10-11-15 is the product $(z_1 z_2 \cdots z_m)$ while b_0 is $(p_1 p_2 \cdots p_n)$. Similarly a_1 is a sum of terms of the form $[(z_1 z_2 \cdots z_{m-1}) + (z_1 z_2 \cdots z_{m-2} z_m) + (z_1 z_2 \cdots z_{m-3} z_{m-1} z_m) + \cdots]$ in which each term contains $m - 1$ zeros. The coefficient b_1 is formed the same way except it involves poles. Eq. 10-11-17 gives

$$K_p' = \frac{K a_0}{b_0 - K a_0} = \frac{K(\text{Product of } z's)}{(\text{Product of } p's) - K(\text{Product of } z's)} \quad (10-11-20)$$

For an infinite $K_p' = K_p$, it is seen that $b_0 = Ka_0$. For this condition $K_v' = K_v$ and

$$-\frac{1}{K_v} = \frac{K}{b_0} \left(a_1 - \frac{b_1 a_0}{b_0} \right) = \frac{a_1}{a_0} \frac{Ka_0}{b_0} - \frac{Ka_0}{b_0} \frac{b_1}{b_0} \quad (10-11-21)$$

$$\frac{1}{K_v} = \frac{b_1}{b_0} - \frac{a_1}{a_0} \quad (10-11-22)$$

From the discussion above, the ratios in Eq. 10-11-22 reduce to the sum of the reciprocals of poles and zeros. Thus

$$\frac{a_1}{a_0} = \frac{z_1 z_2 \cdots z_{m-1}}{z_1 z_2 \cdots z_{m-1} z_m} + \frac{z_1 z_2 \cdots z_{m-2} z_m}{z_1 z_2 \cdots z_{m-2} z_{m-1} z_m} + \cdots \quad (10-11-23)$$

$$\frac{a_1}{a_0} = \frac{1}{z_m} + \frac{1}{z_{m-1}} + \cdots \quad (10-11-24)$$

The ratio b_1/b_0 reduces in a similar manner. Thus

$$\frac{1}{K_v} = \sum_{i=1}^n \frac{1}{p_i} - \sum_{i=1}^m \frac{1}{z_i} \quad (10-11-25)$$

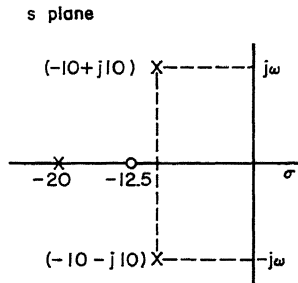


Fig. 10-11-1.

Example 10-11-2. The dynamic error constant relationship to closed-loop poles and zeros is to be studied for a unity feed-back system represented by the poles and zeros of Fig. 10-11-1. From Eq. 10-11-20 the gain constant K for an infinite K_p is found as

$$K = \frac{b_0}{a_0} = \frac{(10 + j10)(10 - j10)(20)}{(12.5)} = 320 \quad (10-11-26)$$

For this value of K

$$\frac{1}{K_v} = \left(\frac{1}{10 - j10} + \frac{1}{10 + j10} + \frac{1}{20} \right) - \left(\frac{1}{12.5} \right) \quad (10-11-27)$$

$$K_v = 14.3 \quad (10-11-28)$$

It is important to notice the effects on K_v of various combinations of poles and zeros. For example, the more poles there are the lower the velocity constant. Poles close to the origin cause low velocity constants unless there are zeros nearby. In contrast zeros, to a certain extent, improve K_v . In Ex. 10-11-2 the zero at -12.5 if moved to -6.67 results in an infinite K_v . The relation in Eq. 10-11-25 is used later in the study of root locus and the choice of closed-loop pole-zero patterns. A relation for K_a can also be established, but it is not as simple or as important as Eq. 10-11-25.

PROBLEMS

10-1. The impulse response of a system is $\varepsilon^{-2t} + \varepsilon^{+4t}$. The time function $r(t) = \cos 5t$ is applied to the system. The output of the system is to be found by using Laplace transforms. The so-called steady-state sinusoidal response is to be found using the methods of Chapters 3 and 6. Explain the results.

10-2. Repeat Prob. 10-1 with $t(t) = \varepsilon^{-2t} + \varepsilon^{-4t}$ and $r(t) = \cos 5t$.

10-3. Convolve ε^{-4t} with $u_{-1}(t)$.

10-4. Convolve the derivative of ε^{-4t} with the integral of $u_{-1}(t)$.

10-5. Convolve the integral of ε^{-4t} with the derivative of $u_{-1}(t)$.

10-6. Convolve ε^{-t} with $u_{-2}(t)$.

10-7. Convolve the derivative of ε^{-t} with the integral of $u_{-2}(t)$.

10-8. Convolve the second integral of ε^{-t} with the second derivative of $u_{-2}(t)$.

10-9. Convolve $1 - \varepsilon^{-t}$ with ε^{-2t} .

10-10. Convolve the derivative of $1 - \varepsilon^{-t}$ with the integral of ε^{-2t} .

10-11. Convolve the second derivative of $1 - \varepsilon^{-t}$ with the second integral of ε^{-2t} .

10-12. Convolve the second integral of $1 - \varepsilon^{-t}$ with the second derivative of ε^{-2t} .

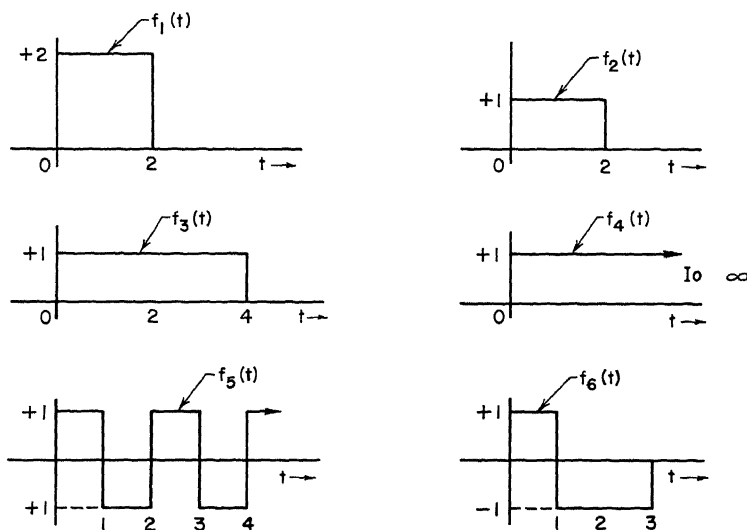


Fig. 10-P-15.

- 10-13.** Convolve $(1 - e^{-2t})$ with $\frac{1}{2}(1 - e^{-t})$.
10-14. Convolve the derivative of $(1 - e^{-2t})$ with the integral of $\frac{1}{2}(1 - e^{-t})$.
10-15. Convolve $f_1(t)$ of Fig. 10-P-15 with itself.
10-16. Convolve $f_5(t)$ of Fig. 10-P-15 with itself.
10-17. Convolve $f_6(t)$ of Fig. 10-P-15 with itself.
10-18. Convolve $f_2(t)$ with $f_3(t)$. Both functions are shown in Fig. 10-P-15.
10-19. Convolve $f_4(t)$ with $f_5(t)$. Both functions are shown in Fig. 10-P-15.
10-20. Convolve $f_3(t)$ with $f_5(t)$. Both functions are shown in Fig. 10-P-15.
10-21. Convolve $f_5(t)$ with $f_6(t)$. Both functions are shown in Fig. 10-P-15.
10-22. Given the transform of a unit-step function response as

$$C(s) = \frac{6(s + 2.5)}{2.5s(s + 2)(s + 3)}$$

Plot the time functions corresponding to

$$C_1(s) = \frac{2}{s(s + 2)} \quad \text{and} \quad C_2(s) = \frac{3}{s(s + 3)}$$

and predict the approximate plot of $c(t)$. Actually calculate $c(t)$ and plot as a check.

10-23. Given

$$T(s) = \frac{6(s + 2.5)}{2.5(s + 2)(s + 3)}$$

Plot the frequency response corresponding to

$$T_1(s) = \frac{2}{s + 2} \quad \text{and} \quad T_2(s) = \frac{3}{s + 3}$$

and predict the approximate frequency plot obtained from $T(s)$. Actually calculate $T(j\omega)$ and plot as a check.

10-24. Given the transform of a unit-step function response as

$$C(s) = \frac{50(s + 3)(s + 6)}{18s(s + 1)(s + 5)(s + 10)}$$

Divide this $C(s)$ into an appropriate $C_1(s)$ and $C_2(s)$ and finish the problem in a manner similar to Prob. 10-22.

10-25. Given

$$T(s) = \frac{50(s + 3)(s + 6)}{18(s + 1)(s + 5)(s + 10)}$$

Divide this $T(s)$ into an appropriate $T_1(s)$ and $T_2(s)$ and finish the problem in a manner similar to Prob. 10-23.

10-26. Prob. 10-22 is repeated with

$$C(s) = \frac{6(s + 1)}{s(s + 2)(s + 3)} \quad ; \quad C_1(s) = \frac{2}{s(s + 2)}$$

and

$$C_2(s) = \frac{3}{s(s+3)}$$

10-27. Prob. 10-23 is represented with

$$T(s) = \frac{6(s+1)}{(s+2)(s+3)} \quad ; \quad T_1(s) = \frac{2}{s+2}$$

and

$$T_2(s) = \frac{3}{s+3}$$

10-28. Predict the general shape of the time response corresponding to

$$C(s) = \frac{10(s+2.1)^2}{(2.1)^2 s(s+2)(s+5)}$$

Next, actually plot the $c(t)$ curve to check the results.

10-29. Predict the general shape of the frequency response corresponding to

$$T(s) = \frac{10(s+2.1)^2}{(2.1)^2 (s+2)(s+5)}$$

Next, actually plot the $T(j\omega)$ curve to check the results.

10-30. From the curves of Fig. 10-7-15a through e, replot on one sheet of paper, the curves with the zero at -0.5 and the pole on the negative real axis taking on the values of -0.5 , -1 , -2 , -5 and $-\infty$.

10-31. Repeat Prob. 10-30 except locate the zero at -2 .

10-32. Repeat Prob. 10-30 except locate the zero at -5 .

10-33. Repeat Prob. 10-30 except locate the zero at $-\infty$.

10-34. Given the transfer function

$$T(s) = \frac{(s+1)(s+2)(s+3)}{(s+0.5)(s^2+s+1)(s+5)}$$

Write all the possible transfer functions that have the same attenuation characteristics and differ only in phase from the $T(s)$ as given.

10-35. In Table 10-10-1, the asymptotes for the attenuation diagrams in transfer functions (b) and (c) are identical. Sketch the actual attenuation curve for both parts to see how they differ.

10-36. Check all diagrams in Table 10-10-1.

10-37. Use Eq. 10-11-20 to find the position constant for a unity feedback system with closed-loop poles at $-20 \pm j20$ and no finite zeros (a_0 is unity in this case). Let $K = 400$. What is K_v for this system? What is K_p ?

10-38. A Type 1 system with closed-loop poles at $-5 \pm j10$ and no finite zeros is to be compensated to increase K , five times. What is the present K and K_v ? A closed-loop pole-zero pair is to be added on the negative real axis near the origin to accomplish the compensation. Which singularity must be nearest the origin? If the zero is located at -0.75 , where must the pole be located?

10-39. Expand the product of factors

$$(s + 1)(s + 2 + j2)(s + 2 - j2)(s + 3)$$

and identify the way the coefficients in the resulting polynomial $s^4 + a_3s^3 + a_2s^2 + a_1s + a_0$ are formed. (Sums of various products of the zeros.)

11

ROOT-LOCUS METHODS

11-1. Introduction. A knowledge of the closed-loop poles and zeros of a control system is necessary for obtaining various responses. The analysis and design methods of Chapters 7, 8, and 9 in this text are concerned mainly with the open-loop poles and zeros and thereby avoid the problem of factoring the system characteristic equation. However, these open-loop methods allow only indirect correlation between closed-loop transient performance and various open-loop system parameters. The root-locus methods introduced in this chapter allow study of a control system while observing through a semigraphical procedure the general relationships between open-loop and closed-loop poles and zeros.

11-2. The Basis of the Direct Root Locus.¹ The basic equation for a unity feedback system like that illustrated in Fig. 11-2-1 is

$$\frac{C(s)}{R(s)} = T(s) = \frac{KG(s)}{1 + KG(s)} \quad (11-2-1)$$

The system gain, K , is shown explicitly, since it enters directly in the root-locus procedures that follow. For the cases in which the transfer function, $KG(s)$, is a ratio of polynomials

$$KG(s) = \frac{q(s)}{p(s)} \quad (11-2-2)$$

$$T(s) = \frac{q(s)}{p(s) + q(s)} = \frac{q(s)}{v(s)} \quad (11-2-3)$$

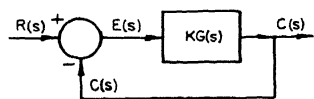


Fig. 11-2-1.

and it is apparent that the closed-loop zeros of $T(s)$ are the same as the open-loop zeros. The closed-loop poles are the zeros of $p(s) + q(s)$ or of $1 + KG(s)$. Normally the procedure involved in determining the closed-loop poles is to factor

$$p(s) + q(s) = 0 \quad (11-2-4)$$

so as to obtain

$$(s + p_1)(s + p_2) \cdots (s + p_n) = 0 \quad (11-2-5)$$

If the open-loop gain K is allowed to vary, the roots of Eq. 11-2-4 vary in some manner. A plot of these roots in the Laplace s -plane as K varies is called the direct root locus. This plot must have as many branches as there are roots of Eq. 11-2-4. If these roots had to be found by factoring Eq. 11-2-4 for each value of K , the root locus would be of no use; however, it is obtained by using the other form for the characteristic equation

$$1 + KG(s) = 0 \quad (11-2-6)$$

or

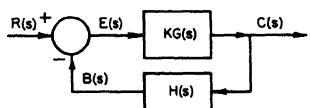
$$KG(s) = -1 \quad (11-2-7)$$

Instead of factoring Eq. 11-2-6, use is made of the fact that $KG(s)$ is a function of a complex variable. Eq. 11-2-7 (and Eq. 11-2-6) specifies two things: first, the magnitude, $|KG(s)|$, must equal 1, and second, the angle of $KG(s)$ must be an odd multiple of 180° . The direct root locus in the s -plane is determined by finding those values of s that satisfy the angle condition above. If the angle condition is satisfied by $s = s_1$, certainly a value of $K = K_1$ can be found to satisfy the magnitude equation where

$$K_1 = \frac{1}{|G(s_1)|} \quad (11-2-8)$$

While the procedure that is used in plotting the direct root locus makes use of the angle criterion, it is not a haphazard search for values of s . This becomes apparent when rules for the construction of the root locus are considered.

The direct root locus can also be applied to control systems of the type shown in Fig. 11-2-2. In this case



$$T(s) = \frac{KG(s)}{1 + KG(s)H(s)} \quad (11-2-9)$$

the characteristic equation can be written as

$$KG(s)H(s) = -1 \quad (11-2-10)$$

Fig. 11-2-2.

It is necessary to consider $KG(s)H(s)$ in this case in the same manner as $KG(s)$ is considered in the unity feedback case. In addition, the poles of $H(s)$ appear as zeros of the closed-loop function $T(s)$.

11-3. The Basis for the Inverse Root Locus. The problem of control system synthesis frequently requires a study of the open-loop poles and zeros when a certain closed-loop pole-zero pattern is specified. This is the reverse of the previous direct root locus process. Since closed-loop performance is generally specified, certain closed-loop poles and zeros are required. In addition, a part of the open-loop system will probably be fixed by certain major components, thereby imposing restrictions on the closed-loop transfer function. It is helpful to be able to visualize these restrictions. The inverse root locus makes such a visualization possible.

Eq. 11-2-1 can be rearranged as

$$KG(s) = \frac{T(s)}{1 - T(s)} \quad (11-3-1)$$

$$\frac{q(s)}{p(s)} = \frac{q(s)}{v(s) - q(s)} \quad (11-3-2)$$

A similarity of form exists between Eqs. 11-2-2 and 11-3-2 with the exception of a sign difference. The open-loop poles in this case are the roots of

$$v(s) - q(s) = 0 \quad (11-3-3)$$

or

$$1 - T(s) = 0 \quad (11-3-4)$$

Again, it is possible, instead of factoring these polynomials, to plot a locus of roots as the gain implicit in $T(s)$ is varied. This locus is called the inverse root locus and the points on it must satisfy the following conditions.

$$|T(s)| = 1 \quad (11-3-5)$$

$$\text{ang } T(s) = 2\pi n \quad ; \quad n = 0, \pm 1, \pm 2, \dots \quad (11-3-6)$$

The inverse root locus is determined by conditions similar to those for the direct root locus except that the angle criterion calls for an even multiple of 180° instead of an odd multiple.

If feedback elements are present so that an $H(s)$ other than unity is used, Eq. 11-2-9 serves as a starting point for developing the inverse equations:

$$T(s) = \frac{1}{H(s)} \frac{KG(s)H(s)}{1 + KG(s)H(s)} \quad (11-3-7)$$

$$KG(s)H(s) = \frac{T(s)H(s)}{1 - T(s)H(s)} \quad (11-3-8)$$

In this case $T(s)H(s)$ must be treated as $T(s)$ is treated in the unity feedback case.

11-4. Rules for the Direct Root Locus. The rules for constructing the direct root locus are developed for unity feedback cases. For other cases $KG(s)H(s)$ is used in place of $KG(s)$ in all manipulations. In establishing the angle of $KG(s)$ for a given value of s , it is convenient to consider the factored form

$$KG(s) = \frac{K(s + z_1)}{s(s + p_1)(s + p_2)} \quad (11-4-1)$$

Fig. 11-4-1a shows the zeros and poles of this open-loop function. For a particular value of $s = s_1$, the factors $s_1 + z_1$, s_1 , $s_1 + p_1$, $s_1 + p_2$ can be represented as phasors as in Fig. 11-4-1b. The important point to ob-

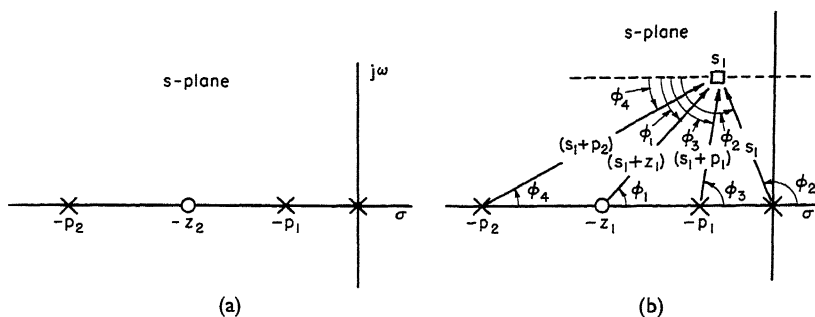


Fig. 11-4-1.

serve in Fig. 11-4-1b is that *all phasor angles can be measured with a protractor at the s_1 point being considered* instead of moving the protractor to $-z_1$, $-p_1$, $-p_2$, etc., in succession. The problem in plotting the direct root locus is to find values of s that cause the net angle of $KG(s)$ to be an odd multiple of 180° . From Fig. 11-4-1b this requires

$$\varphi_1 - (\varphi_2 + \varphi_3 + \varphi_4) = (2j + 1)\pi \quad ; \quad j = 0, \pm 1, \pm 2, \dots \quad (11-4-2)$$

The following rules help to find values of s that satisfy such an angle equation.

RULE 1. REAL AXIS SEGMENTS. Certain segments of the real axis are generally parts of the root locus. Since the angle of a factor such as $s + z_1$ will be either 180° or 0° for real values of s and z_1 , it can be seen that the segments to be used are to the left of an odd number of poles and zeros. It must be added that conjugate pairs of complex poles contribute a net angle of zero to the angle of $KG(s)$ for values of s along the real axis.

Example 11-4-1. The open-loop transfer function

$$KG(s) = \frac{K(s + 10)}{s(s + 25)(s + 40)(s^2 + 10s + 50)} \quad (11-4-3)$$

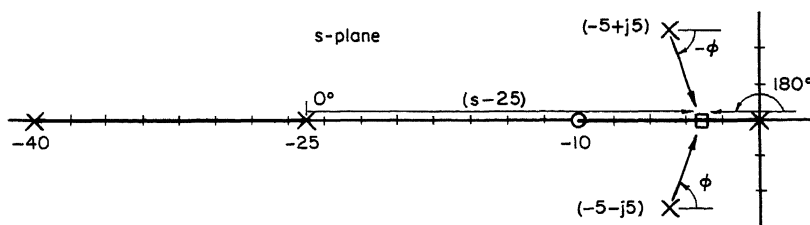


Fig. 11-4-2.

has its poles and zero plotted in the s -plane of Fig. 11-4-2. The real axis segments of a unity-feedback-system direct root locus for Eq. 11-4-3 also are shown in Fig. 11-4-2. The factors $s + 10$, $s + 25$, $s + 40$ all have an angle of zero for a value of s between 0 and -10 while the factor s has an angle of 180° . The factors $s + 5 + j5$ and $s + 5 - j5$ can be seen to contribute equal but opposite angles $\pm\varphi$. Thus any s on the real axis segment from 0 to -10 gives

$$\text{ang } KG(s) = 0^\circ - (180^\circ + 0^\circ + 0^\circ + \varphi - \varphi) = -180^\circ \quad (11-4-4)$$

and therefore forms a part of the direct root locus. This means that for certain values of K the points on this segment are roots of

$$1 + KG(s) = 0 \quad (11-4-5)$$

Similarly the segment from -25 to -40 , which is to the left of 3 real poles and zeros, is part of the locus since

$$\text{ang } KG(s) = 180^\circ - (180^\circ + 180^\circ + 0 + \varphi - \varphi) = -180^\circ \quad (11-4-6)$$

Thus any segment of the real axis to the left of an odd number of poles and zeros forms a part of the locus.

RULE 2. STARTING AND ENDING POINTS. As stated before, the direct root locus gives the closed-loop poles as the open-loop gain K is varied. For $K = 0$, it is necessary that $G(s) \rightarrow \infty$ or approach a pole in order for the magnitude requirement

$$|KG(s)| = 1 \quad (11-4-7)$$

to be satisfied. Alternatively, from

$$1 + KG(s) = 0 = p(s) + q(s) \quad (11-4-8)$$

as $K \rightarrow 0$, $g(s) \rightarrow 0$ and this leaves $p(s) = 0$; but $p(s)$ contains simply the factors that give the open-loop poles. This means that the closed-loop poles degenerate to the open-loop poles as the open-loop gain K goes to zero.

If $K \rightarrow \infty$, it is necessary that $G(s)$ approach zero in order to satisfy the magnitude criterion. From Eq. 11-4-8

$$\frac{1}{K} + G(s) = 0 \quad (11-4-9)$$

and if $K \rightarrow \infty$, this degenerates to

$$G(s) = 0 \quad (11-4-10)$$

or to the open-loop zeros, including those at infinity. Thus, as $K \rightarrow \infty$, the closed-loop poles move toward the open-loop zeros. The above facts are stated concisely as Rule 2: *As the open-loop gain K is increased from zero to infinity, the branches of the locus of the closed-loop poles start at open-loop poles and terminate on open-loop zeros (including those at infinity).*

Example 11-4-2. A unity feedback system with

$$KG(s) = \frac{K(s+1)}{s} \quad (11-4-11)$$

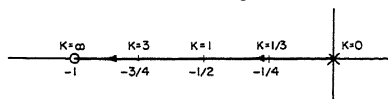


Fig. 11-4-3.

is considered in this example so that a very simple root locus results. As shown in Fig. 11-4-3, the section of the real axis from 0 to -1 forms all of the root locus. The actual closed-loop characteristic equation is

$$1 + \frac{K(s+1)}{s} = 0 \Rightarrow s + K(s+1) = 0 \quad (11-4-12)$$

This gives the closed-loop pole as

$$s = \frac{-K}{1+K} \quad (11-4-13)$$

and also

$$K = \frac{-s}{s+1} \quad (11-4-14)$$

From Eq. 11-4-13, the correctness of Rule 2 in this case can be established since for $K=0$, $s=0$ and for $K \rightarrow \infty$, $s \rightarrow -1$. Between these extreme values, Eq. 11-4-14 can be used to determine the values of K that give a closed-loop pole at a specific value of s on the indicated locus.

Example 11-4-3. As a second example of Rule 2

$$KG(s) = \frac{K}{s} \quad (11-4-15)$$

is studied. This transfer function has a single zero at infinity as well as a pole at the origin. The closed-loop characteristic equation is now

$$1 + \frac{K}{s} = \frac{s+K}{s} = 0 \quad (11-4-16)$$

$$s = -K \quad (11-4-17)$$

From Eq. 11-4-17 it can be seen that the closed-loop pole moves from $s=0$ to $s=-\infty$ along the negative real axis as K varies from 0 to ∞ . Since most open-loop transfer functions have one or more zeros at infinity, there will generally be branches of the direct root locus going to infinity along various paths. The following rule considers such situations.

RULE 3. ASYMPTOTES FOR LARGE s . When there are zeros or poles in $KG(s)$ at infinity, asymptotes play an important role in the direct root locus plots. The equation

$$\frac{K}{s^n} = -1 \quad (11-4-18)$$

which has n zeros at infinity is considered. The roots of the equation, $s^n + K = 0$, formed from Eq. 11-4-18, have varying magnitudes as K varies, but all roots follow a set of lines emanating from the origin of the s -plane. These lines make angles of $(2k + 1)\pi/n$ radians ($k = 0, 1, \dots, n - 1$), with the positive real axis. For example, if $n = 3$ so that $s^3 + K = 0$ is the equation under consideration, the roots follow paths as indicated in Fig. 11-4-4a. If the equation is of the form

$$\frac{K}{(s + c)^n} = -1 \quad (11-4-19)$$

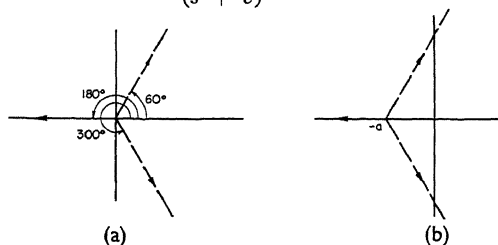


Fig. 11-4-4.

the root loci emanate from $s = -c$, but still make the same angles in the s -plane as illustrated for $n = 3$ in Fig. 11-4-4b.

The characteristic equation can generally be written in the form

$$KG(s) = \frac{K(s^m + a_{m-1}s^{m-1} + \dots + a_0)}{s^n + b_{n-1}s^{n-1} + \dots + b_0} = -1 \quad (11-4-20)$$

with $n > m$. For very large values of s , only the higher order terms need be considered, so that Eq. 11-4-20 can be approximated as

$$KG(s) \approx \frac{Ks^m}{s^n} = \frac{K}{s^{n-m}} = -1 \quad (11-4-21)$$

Eq. 11-4-21 indicates that the root locus has $n - m$ asymptotic branches which are approached at large values of s . These asymptotes make angles of $(2k + 1)\pi/(n - m)$ radians ($k = 0, 1, 2, \dots, n - m - 1$), with the positive real s -plane axis.

Since most work in plotting root loci involves a small portion of the s -plane near the origin, it is necessary to be able to determine a suitable point of intersection for the asymptotes. Neglecting all but the highest power of s as in Eq. 11-4-21 indicates that from infinity the asymptotes appear to come together at the origin in the s -plane. However, as the finite region nearer the origin but still for relatively large s is considered, more powers of s in Eq. 11-4-20 must be retained. In order to use the $n - m$ asymptotes established above while obtaining a better approximation to the actual Eq. 11-4-20, a relation of the form

$$KG(s) \approx \frac{K}{(s + c)^{n-m}} = -1 \quad (11-4-22)$$

is generally used. The constant c is chosen so that the s^{n-m-1} term in the expansion of 11-4-22 will match the similar term in an expansion of Eq. 11-4-20 in the form

$$KG(s) = \frac{K}{s^{n-m} + (b_{n-1} - a_{m-1})s^{n-m-1} + \dots} = -1 \quad (11-4-23)$$

This requires that $(n - m)c = b_{n-1} - a_{m-1}$. From earlier discussion the term b_{n-1} is known to be the negative of the sum of the open-loop poles while a_{m-1} is the negative of the sum of the open-loop zeros. Since Eq. 11-4-22 has $n - m$ root loci originating at $-c$, the root locus asymptotes for the actual system originate at

$$-c = \frac{-(b_{n-1} - a_{m-1})}{n - m} \quad (11-4-24)$$

$$-c = \frac{\Sigma(\text{Open loop poles}) - \Sigma(\text{Open loop zeros})}{n - m} \quad (11-4-25)$$

This scheme determines an origin for $n - m$ intersecting asymptotes making the specified angles with the positive real axis.

Finally, Rule 3 can be stated as follows. *The segments of the root locus for large s and $n > m$ can be approximated by $n - m$ straight-line asymptotes making angles of*

$$\frac{(2k + 1)\pi}{n - m} ; \quad k = 0, 1, 2, \dots, n - m - 1 \quad (11-4-26)$$

with the positive real axis and intersecting at the point given by

$$\frac{\frac{\sum_1^n (\text{Open loop poles}) - \sum_1^m (\text{Open loop zeros})}{1}}{n - m} \quad (11-4-27)$$

where n is the number of finite poles and m is the number of finite zeros in the transfer function $KG(s)$.

Example 11-4-4. Rule 3 is applied to the transfer function of Example 11-4-1, namely

$$KG(s) = \frac{K(s + 10)}{s(s + 25)(s + 40)(s^2 + 10s + 50)} \quad (11-4-28)$$

In this case $n = 5$, $m = 1$, and therefore there are four asymptotes, which make

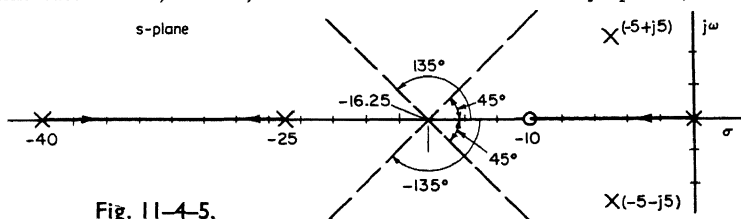


Fig. 11-4-5.

angles of $\pi/4$, $3\pi/4$, $5\pi/4$, and $7\pi/4$ radians with the positive real axis, as shown in Fig. 11-4-5. The asymptotes intersect at

$$\frac{[0 - 25 - 40 + (-5 + j5) + (-5 - j5)] - [-10]}{4} = -16.25 \quad (11-4-29)$$

The real axis segments are also shown with arrows to indicate the direction of movement of the closed-loop poles as the open-loop gain K is increased. As mentioned before, the closed-loop poles start at the open-loop poles and move toward the open-loop zeros, of which four are at infinity in this case. The straight-line asymptotes indicate how these four zeros at infinity are approached.

It should be noted that the segment from -25 to -40 is also a part of the root locus and that two closed-loop poles must move toward each other. At some value of gain they coincide, giving a double closed-loop pole. It is found that the locus then leaves the real axis with two complex segments. There also must be segments starting at the two complex poles at $-5 \pm j5$.

RULE 4. COMPLEX CONJUGATE SEGMENTS. Since the closed-loop characteristic equation $1 + KG(s) = 0$, involves only real coefficients of terms in s , *all closed-loop complex roots must occur in conjugate pairs*. As stated previously, there must be as many branches as there are closed-loop poles.

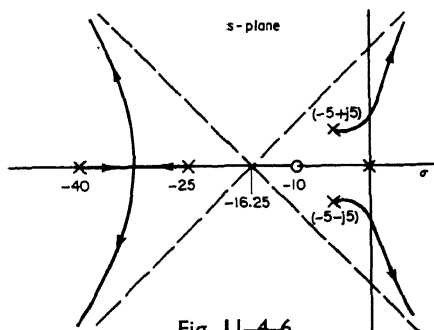


Fig. 11-4-6.

Example 11-4-5. For the $KG(s)$ of Example 11-4-4 there are four complex branches as shown in Fig. 11-4-6. Several factors must be considered in order to sketch these branches. The first of these factors is the point of departure of the two segments between $s = -25$ and -40 . The second of these is the angle at which the complex segments leave the complex poles at $-5 \pm j5$. A process of trial and error with a protractor could be used since all that is needed are the values of s that make the angle of $KG(s)$ equal an odd multiple of 180° . However, later rules establish some simpler methods that are useful in many cases.

Five segments of the root locus of Eq. 11-4-28 are now available, giving the movement of five closed-loop poles as K varies. The need for five segments comes from the fact that a fifth-degree polynomial results for the characteristic equation:

$$s^5 + 15s^4 + 1,700s^3 + 13,250s^2 + (50,000 + K)s + 10K = 0 \quad (11-4-30)$$

What has really been sketched in Fig. 11-4-6 are the paths of the roots of Eq. 11-4-30 as K varies from zero to infinity. No factoring has been done; however, it might be pointed out that plotting the exact locus and determining its calibration for specific values of K corresponding to closed-loop poles at various points requires considerable additional work.

RULE 5. DEPARTURE OF COMPLEX SEGMENTS FROM REAL AXIS. As indicated in the previous example, there is a need for determining the point of departure of a complex segment of the root locus from the real axis. Fig. 11-4-7 illustrates a case in which the real axis segment to the

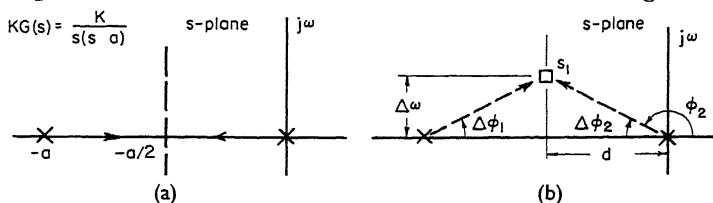


Fig. 11-4-7.

left of an odd number of poles has been indicated as a part of the direct root locus. Since two closed-loop poles approach each other for relatively small values of K and since there is a double zero at infinity requiring asymptotes at $\pm 90^\circ$ emanating from $-a/2$, there must be two complex segments breaking away from the real axis at some point. To locate this point, a test point is chosen just off the real axis in the general vicinity of breakaway, as shown in Fig. 11-4-7b. If the test point is on the locus, the angle of $KG(s)$ must still be an odd multiple of 180° . Thus $-\varphi_2 - \Delta\varphi_1 = \pm 180^\circ$ (since this is the angle for points on the real axis segment and no sudden jump in angle should occur). However, $\varphi_2 = 180^\circ - \Delta\varphi_2$ and therefore

$$-180^\circ + \Delta\varphi_2 - \Delta\varphi_1 = \pm 180^\circ \quad (11-4-31)$$

or

$$\Delta\varphi_1 = \Delta\varphi_2 \quad (11-4-32)$$

In reality, this states that no change should occur in the angle of $KG(s)$ as s leaves the real axis segment of the locus. By using d and $\Delta\omega$ Eq. 11-4-32 can be written as

$$\tan^{-1}\left(\frac{\Delta\omega}{d}\right) - \tan^{-1}\left[\frac{\Delta\omega}{(a-d)}\right] = 0 \quad (11-4-33)$$

For very small departures from the real axis, the first-term approximation from a series expansion for the arctangent function gives

$$\frac{\Delta\omega}{d} - \frac{\Delta\omega}{a-d} = 0 \quad (11-4-34)$$

$$\frac{1}{d} - \frac{1}{a-d} = 0 \quad (11-4-35)$$

The point of departure is now found from Eq. 11-4-35 as $d = a/2$. This value could have been arrived at by direct reasoning in this simple case since only for an isosceles triangle will the two angles $\Delta\varphi_1$ and $\Delta\varphi_2$ be the same. The angles need to be the same here, since only two angles are involved, thus requiring one to be decreased by the same amount as the other is increased to prevent a change in the net angle of $KG(s)$. The

complete root locus in this case follows the real axis to the point $-a/2$ and then follows the asymptotes to infinity.

An equation of the form of Eq. 11-4-33 can always be set up for finding d ; however, as the system grows in complexity, a higher-order equation in d results and the advantages of such a scheme become questionable. When complex poles are introduced, a more involved quadratic term results for each pair of complex poles. These statements are illustrated in the examples that follow. First, in summary, Rule 5 can be stated as follows: *By assuming a test point near the real axis breakaway, setting the change in $\arg KG(s)$ equal to zero and using a linear approximation to the arctangent function, an equation giving the breakaway point can be obtained.*

Example 11-4-6. For

$$KG(s) = \frac{K(s + 10)}{s(s + 4)} \quad (11-4-36)$$

the procedure for applying Rule 5 is illustrated in Fig. 11-4-8a. It should be noted that increases in the angles for zeros have the same effect as decreases in the angles for poles

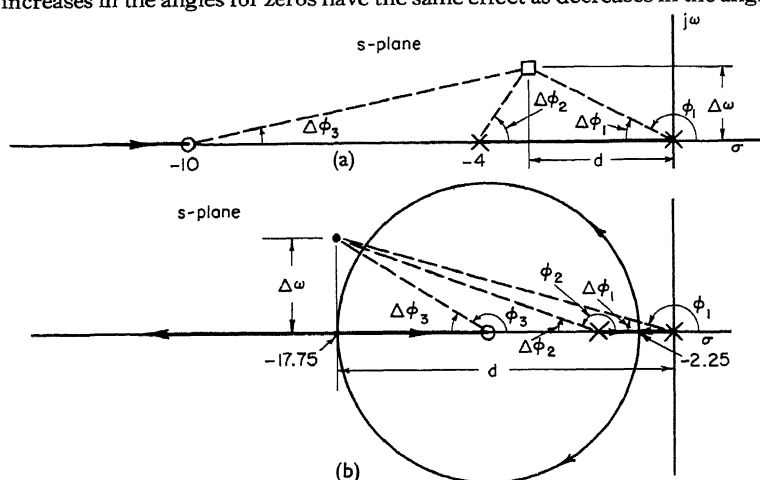


Fig. 11-4-8.

poles. Thus the angle $\Delta\phi_2$ should be the largest angle and in fact it must equal the sum of $\Delta\phi_3$ and $\Delta\phi_1$:

$$\Delta\phi_2 = \Delta\phi_1 + \Delta\phi_3 \quad (11-4-37)$$

$$\tan^{-1}\left(\frac{\Delta\omega}{4-d}\right) = \tan^{-1}\left(\frac{\Delta\omega}{d}\right) + \tan^{-1}\left(\frac{\Delta\omega}{10-d}\right) \quad (11-4-38)$$

The sides of the triangles are set up as normally positive quantities such as $4-d$ and $10-d$. Finally

$$\frac{1}{4-d} \approx \frac{1}{d} + \frac{1}{10-d} \quad (11-4-39)$$

$$d^2 - 20d + 40 = 0 \quad ; \quad d = 2.26, 17.74 \quad (11-4-40)$$

The smaller value of d is logical but the larger value brings up a question as to its usefulness. However, only one zero exists at infinity and the real axis segment

from -10 to $-\infty$ is a part of the locus. Apparently the two complex branches that split off at -2.26 must come back together somewhere outside -10 and proceed along the real axis to the zero at -10 and to the zero at $-\infty$. It would seem reasonable to expect the -17.74 solution above to be the point of return. This fact can be checked by setting up the "no change in angle" relation outside -10 using Fig. 11-4-8b.

$$(-\Delta\varphi_3) - (-\Delta\varphi_2 - \Delta\varphi_1) = 0 \quad (11-4-41)$$

all $\Delta\varphi$'s are decreases in the net angle of each term from the values of s on the real axis. Thus

$$\frac{1}{d-10} \approx \frac{1}{d-4} + \frac{1}{d} \quad (11-4-42)$$

or

$$\frac{1}{4-d} \approx \frac{1}{d} + \frac{1}{10-d} \quad (11-4-43)$$

which is exactly the same as used before. The value $d = 17.74$ does give the point of return of the locus. The rest of the root locus in Fig. 11-4-8b is roughly sketched for completeness.

Example 11-4-7. As a second example of Rule 5 suppose

$$KG(s) = \frac{K}{s(s+4)(s+6)(s+10)} \quad (11-4-44)$$

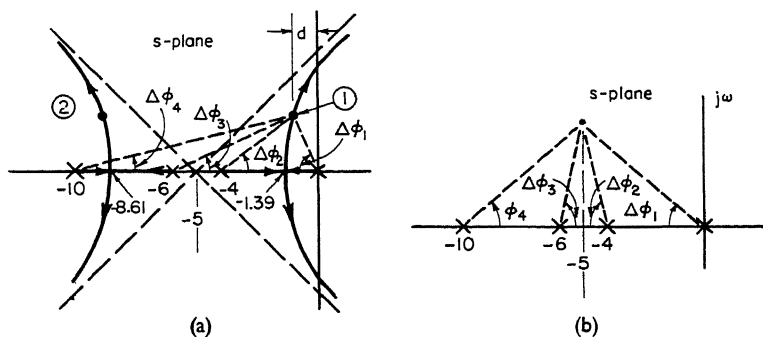


Fig. 11-4-9.

Fig. 11-4-9a illustrates the preliminary information obtained by applying Rules 1 and 3 and the information required to set up the point of departure relationship of Rule 5. For point ①

$$-(\Delta\varphi_2 + \Delta\varphi_3 + \Delta\varphi_4 - \Delta\varphi_1) = 0 \quad (11-4-45)$$

$$\frac{1}{4-d} + \frac{1}{6-d} + \frac{1}{10-d} - \frac{1}{d} = 0 \quad (11-4-46)$$

$$4d^3 - 60d^2 + 248d - 240 = 0 \quad (11-4-47)$$

$$d = 5, 8.61, 1.39 \quad (11-4-48)$$

The roots $d = 1.39$ and $d = 8.61$ are expected, since there are two points of break-away, one between 0 and -4 and the other between -6 and -10 . The root, $d = 8.61$, can be checked by setting up an equation at point ②. However, the presence of the value $d = 5$ is puzzling at first. The relationship that is set up requires no change in angle at the test point from what the angle would be on the real axis. The region from -4 to -6 is considered now. Obviously this is not part

of the root locus, since the angle of $KG(s)$ for s on this segment is -360° . However, if a path left the real axis vertically at $s = -5$, the angle of $KG(s)$ would not change from the 360° value on the segment between -4 and -6 . This is true since two pairs of equal angles are involved as indicated in Fig. 11-4-9b. The angles $\Delta\varphi_3$ and $\Delta\varphi_4$ are increases in the angles of the open-loop pole terms $s + 6$, and $s + 10$ while $\Delta\varphi_1$ and $\Delta\varphi_2$ are decreases in the angles of the open-loop pole terms s and $s + 4$. Since $\Delta\varphi_1 = \Delta\varphi_4$ and $\Delta\varphi_2 = \Delta\varphi_3$, there is no net change in the angle of $KG(s)$, as mentioned above. Therefore, not all solutions of the departure-point equation give actual root-locus departure or arrival points. The increasing order of the equation in d should be observed also.

Example 11-4-8. In order to introduce the effect of complex open-loop poles on the procedure of Rule 5

$$KG(s) = \frac{K}{s(s+10)(s^2+4s+104)} \quad (11-4-49)$$

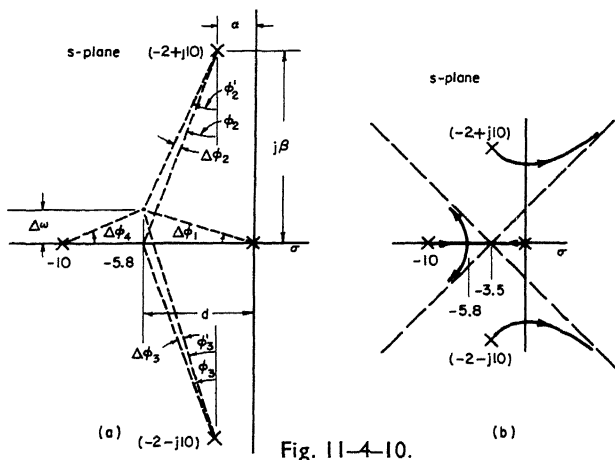


Fig. 11-4-10.

is considered next. The application of Rules 1 and 3 give the information sketched in Fig. 11-4-10a. The real poles present no new problems and so attention is focused on angles $\Delta\varphi_2$ and $\Delta\varphi_3$. Since $\Delta\varphi_2 = \varphi'_2 - \varphi_2$ in the above figure

$$\Delta\varphi_2 = \tan^{-1} \left[\frac{d-a}{\beta - \Delta\omega} \right] - \tan^{-1} \left[\frac{d-a}{\beta} \right] \quad (11-4-50)$$

$$= \tan^{-1} \left[\frac{(d-a)/(\beta - \Delta\omega) - (d-a)/\beta}{1 + [(d-a)/(\beta - \Delta\omega)][(d-a)/\beta]} \right] \quad (11-4-51)$$

$$\approx \tan^{-1} \left[\frac{\Delta\omega(d-a)}{\beta^2 + (d-a)^2} \right] \quad (11-4-52)$$

$\Delta\varphi_3$ can be shown to be approximately the same as $\Delta\varphi_2$ and so the net change in angle due to the complex poles in Fig. 11-4-10a is a decrease of approximately

$$2 \tan^{-1} \frac{\Delta\omega(d-a)}{\beta^2 + (d-a)^2} \quad (11-4-53)$$

The complete equation for d is therefore

$$\Delta\varphi_4 = \Delta\varphi_2 + \Delta\varphi_3 + \Delta\varphi_1 \quad (11-4-54)$$

$$\frac{1}{10-d} = \frac{2(d-2)}{10^2 + (d-2)^2} + \frac{1}{d} \quad (11-4-55)$$

$$d = 5.8, 2.35 \pm j6.27 \quad (11-4-56)$$

Since only real solutions of d are of interest, the value $d = 5.8$ gives the desired breakaway point. Again, a cubic equation has to be factored to obtain this value, making it apparent that this approach is of limited value if a larger number of poles and zeros is involved. In some cases poles that are far away from the point of breakaway can be neglected, thus simplifying the equation in d while retaining reasonable accuracy. The remainder of the root locus is sketched in Fig. 11-4-10b using Rule 6 which is yet to be discussed.

RULE 6. ANGLES OF DEPARTURE AND ARRIVAL AT COMPLEX POLES AND ZEROS. Whenever complex poles and zeros exist in the open-loop transfer function, there are branches of the root locus that leave and arrive at these points. In many cases an angle of departure or arrival is very useful in establishing the general shape of the root locus near these singularities. In fact, it has already been used several times previously in roughing in the root locus.

The desired angle is generally determined as follows for complex poles. For points very close to the complex poles the angles of all terms, except the one for the pole under consideration, are essentially constant at the value computed for s equal to the pole value. In reality, s is not equal to the pole value except for $K = 0$ and so an additional angle for the pole term must be considered. This is the desired but unknown angle. In the original angle criterion for the direct root locus this angle is taken as an unknown pole angle φ . Since it is the only unknown in the angle equation, the solution for φ is possible.

If the angles of arrival at a zero is under consideration, the procedure is exactly the same except that the unknown angle results from a zero term. The root locus enters or leaves the zero or pole under consideration tangent to the line determined by the angle found from the above procedure. In summary, Rule 6 can be stated as follows: *The angles of phasors from all poles and zeros to the one at which the angle of departure or arrival is desired are computed. The root locus angle equation using these angles plus an unknown one for the angle of departure or arrival is written. This angle equation is solved for the unknown angle.*

Example 11-4-9. In order to fill in previous examples, the open-loop transfer function from Examples 11-3-1, 11-3-4, and 11-3-5 is considered again

$$KG(s) = \frac{K(s+10)}{s(s+25)(s+40)(s^2+10s+50)} \quad (11-4-57)$$

with poles and zeros as in Fig. 11-4-11. From the geometry of the figure the angles are computed as listed on the figure or they can be computed from $KG(s)$ by substituting $s = -5 + j5$ in all terms but the pole term $s + 5 - j5$. For a point, s_1 , on the root locus but just off the point $s = -5 + j5$, the angles φ_1 through φ_5 remain ap-

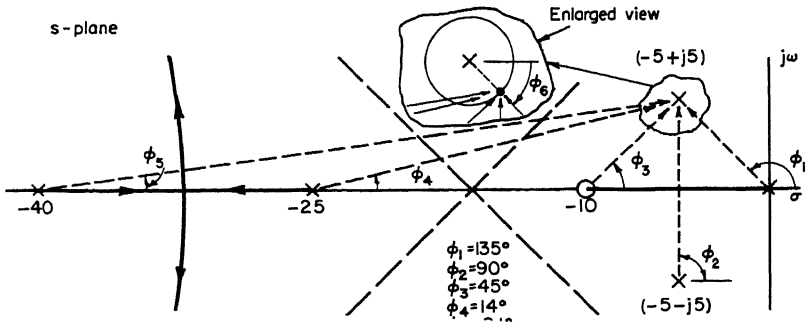


Fig. 11-4-11.

proximately the same but the term $s_1 + 5 - j5$, now has an unknown angle φ_6 . To find φ_6 the root-locus angle criterion gives

$$\varphi_3 - (\varphi_1 + \varphi_2 + \varphi_4 + \varphi_5 + \varphi_6) = 180^\circ \quad (11-4-58)$$

Actually any odd multiple of 180° can be used.

$$45^\circ - (135^\circ + 90^\circ + 14^\circ + 8.1^\circ + \varphi_6) = 180^\circ \quad (11-4-59)$$

$$\varphi_6 = -382.1^\circ = -22.1^\circ \quad (11-4-60)$$

By using this angle and an approximate breakaway point computed by using Rule 5, the sketch of the root locus in Fig. 11-4-12 is developed.

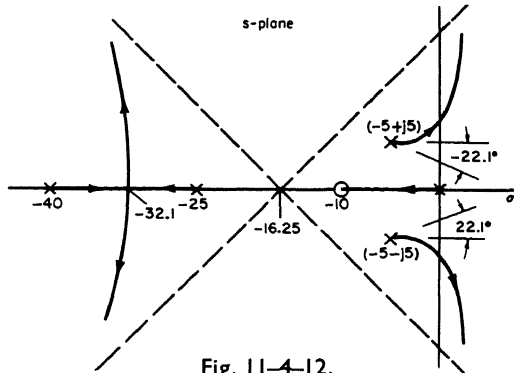


Fig. 11-4-12.

Example 11-4-10. In order to illustrate the angle-of-arrival part of Rule 6, the $KG(s)$ of Eq. 11-4-61 is considered.

$$KG(s) = \frac{K(s^2 + 20s + 200)}{s(s + 2)} \quad (11-4-61)$$

Application of Rules 1, 3, 5, and 6 gives the information sketched in Fig. 11-4-13a. From Rule 6

$$(\varphi_1 + 90^\circ) - (135^\circ + 128.6^\circ) = 180^\circ \quad (11-4-62)$$

$$\varphi_1 = 353.6^\circ = -6.4^\circ \quad (11-4-63)$$

For the point of departure from the real axis between 0 and -2, Rule 5 gives

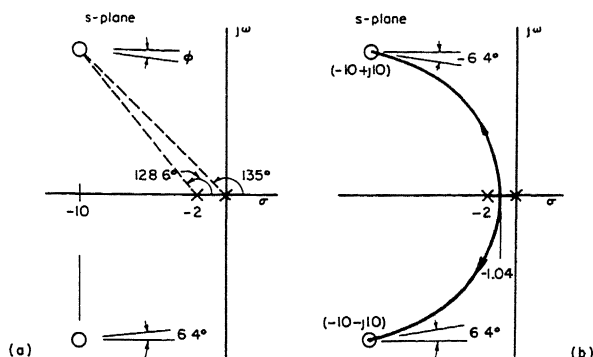


Fig. 11-4-13.

$$\frac{2(10-d)}{(10-d)^2 + 10^2} = \frac{1}{2-d} - \frac{1}{d} \quad (11-4-64)$$

$$9d^2 - 200d + 200 = 0 \quad (11-4-65)$$

$$d = 1.04, 21.2 \quad (11-4-66)$$

the value $d = 1.04$ being the desired departure point. The root locus can now be sketched as in Fig. 11-3-13b.

RULE 7. CALIBRATION OF THE DIRECT ROOT LOCUS. After the various branches of the root locus have been located, it is necessary to calibrate the locus in terms of specific values of the open-loop gain K . This process establishes points that satisfy the magnitude relation

$$|KG(s)| = 1 \quad (11-4-67)$$

obtained from the characteristic equation. Written another way this relation gives the value of K for a specific value of $s = s_1$ on the root locus.

$$K = \frac{1}{|G(s_1)|} \quad (11-4-68)$$

The magnitude $|G(s_1)|$ can be determined by measurement of lengths graphically or by analytical calculation. The graphical procedure assumes an s -plane plot with equal scales on the σ - and $j\omega$ -axes.

In order to illustrate the procedure, the transfer function

$$KG(s) = \frac{K(s+b)}{s(s+a)} \quad (11-4-69)$$

having the root locus in Fig. 11-4-14, is studied. In finding the value of K that gives closed-loop poles at the points labeled ① on the root locus, Eq. 11-4-68 can be used as follows:

$$K = \frac{1}{d_3/d_1d_2} = \frac{d_1d_2}{d_3} \quad (11-4-70)$$

The lengths l_1 , l_2 , and l_3 can be measured with a pair of dividers and checked against the σ - or $j\omega$ -scales to give d_1 , d_2 , and d_3 . A few such calculations generally provide adequate calibration of the root locus.

While there are other special methods of calibrating certain points on the root locus, the use of Eq. 11-4-68 is fastest for setting K at a number of general points. Rule 7 can then be summarized as follows: *To set the value of K at points on the root locus, the relation*

$$K = \frac{1}{|G(s)|} \quad (11-4-71)$$

is used graphically.

Example 11-4-11. As an example of the use of Rule 7, the following $KG(s)$ is considered:

$$KG(s) = \frac{K}{s(s+4)} \quad (11-4-72)$$

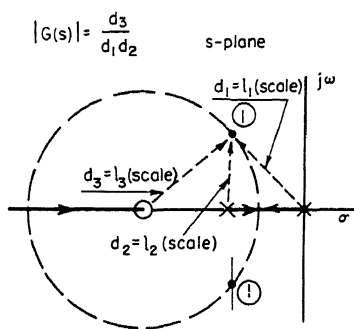


Fig. 11-4-14.

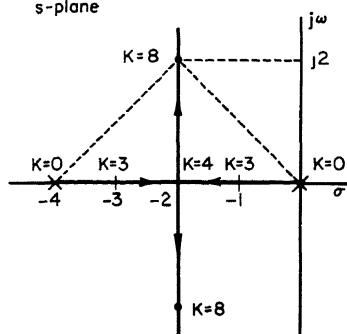


Fig. 11-4-15.

The root locus for this function when used in a unity feedback closed-loop system is shown in Fig. 11-4-15. To find the gain corresponding to the point (-1)

$$K = (1)(3) = 3 \quad (11-4-73)$$

At -2

$$K = (2)(2) = 4 \quad (11-4-74)$$

while at $-2 \pm j2$

$$K = (2\sqrt{2})(2\sqrt{2}) = 8 \quad (11-4-75)$$

It might be pointed out here that when two or more zeros exist at infinity in the transfer function, $KG(s)$, the sum of the closed-loop poles is constant. This is apparent in this case since

$$1 + KG(s) = 0 = s^2 + 4s + K \quad (11-4-76)$$

where -4 is the sum of poles.

This means that as one or more poles move left with K changing, the others must move right to keep the sum constant. In this case the constant sum of the poles

provides a means of pairing points on the horizontal axis that result in the same K . Thus -1 and -3 are such a pair.

Another useful item occurs when one or more open-loop poles exist at the origin as in this example. The product of the closed-loop poles is then proportional to K , as can be seen in Eq. 11-4-76 above, where the product of closed-loop poles is equal to K . Thus for closed-loop poles at $-2 \pm j2$, the product of poles gives

$$K = (-2 + j2)(-2 - j2) = 8 \quad (11-4-77)$$

11-5. General Comments on Construction of the Direct Root

Locus. The preceding section has set forth some very useful rules for constructing and calibrating the direct root locus. However, one important fact should be emphasized before continuing. The construction of root loci is not all done analytically, as the rules might lead one to think. The complex sections of the root locus have to be refined graphically! A rough sketch can be made fairly easily, but to obtain good accuracy a protractor or some such device² has to be used to check the angle of $KG(s)$ at proposed locus points which are not determined analytically. If a point on the rough sketch does not give an angle that is an odd multiple of 180° , the test point is moved and rechecked until it does. The method illustrated in Fig. 11-4-1b should be used in measuring angles. In moving a test point, it is helpful to notice the effect on the angles of the zeros and poles, as compared with previous resultant angles, in order to determine the direction in which to move a point.

Since the imaginary axis is the dividing line between poles that lead to stable and unstable response modes, an analytical method of locating the points, where the loci cross the $j\omega$ -axis as well as determining the corresponding gain is useful. By applying Routh's test to the closed-loop characteristic equation, the above data can be easily obtained in simple cases. The procedure is best illustrated by an example.

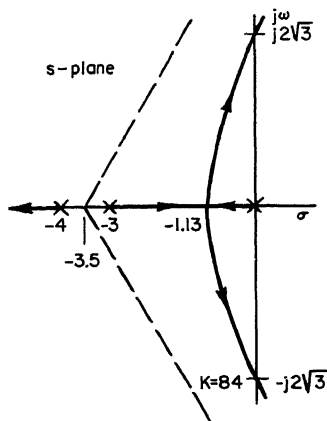


Fig. 11-5-1.

² A device called a Spirule is a useful tool for constructing root loci (The Spirule Co., Whittier, California).

Example 11-5-1. For this example a unity feedback system with

$$KG(s) = \frac{K}{s(s+3)(s+4)} \quad (11-5-1)$$

is used. The three zeros at infinity mean asymptotes exist at $\pm 60^\circ$ and 180° centered at -3.5 as shown in Fig. 11-5-1. The breakaway point is found at -1.13 . The characteristic equation can be written

$$s^3 + 7s^2 + 12s + K = 0 \quad (11-5-2)$$

for which the Routh table is

s^3	1	12
s^2	7	K
s^1	$84 - K$	
s^0	K	

In order to have an imaginary axis crossing a row of terms must be zero. The first such row would be the third with $K = 84$. This means that for this K value, $7s^2 + 84$ is a factor of the characteristic equation, or $s = \pm j2\sqrt{3}$ are the crossing points. Since the sum of roots in this case is constant at -7 , the point $s = -7$ also corresponds to $K = 84$ as is checked by Rule 7

$$K = (7)(4)(3) \quad (11-5-3)$$

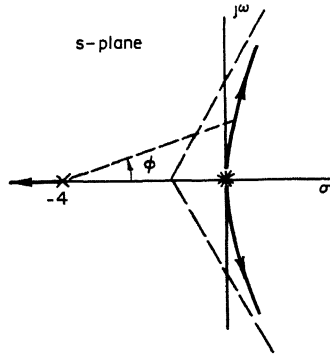


Fig. 11-5-2.

Example 11-5-2. Another concept that is frequently useful is the behavior of the root locus for a system with a double pole at the origin. A unity feedback system with

$$KG(s) = \frac{K}{s^2(s+4)} \quad (11-5-4)$$

is considered. The pole pattern in Fig. 11-5-2 illustrates why it is necessary for the root locus to leave the double pole vertically. For very small movements vertically each pole at the origin contributes an angle of 90° while the pole at -4 contributes 0° . As the locus moves away it must move into the right-half s -plane so that the total angle due to the double pole decreases as much from 180° as the angle due to the pole at -4 increases. Another way of approaching this example is to consider that for very small values of s the $s+4$ factor is approximately 4 leaving

$$KG(s) \approx \frac{K}{4s^2} \quad (11-5-5)$$

The characteristic equation becomes

$$\frac{K}{4s^2} = -1 \quad (11-5-6)$$

for which s must be purely imaginary, or in other words the root locus must leave the origin along the imaginary axis.

11-6. The Inverse Root Locus Rules.^{3, 4} As mentioned in Art. 11-3, the reverse relation between open-loop and closed-loop singularities can be used instead of the direct root locus of Arts. 11-2 and 11-5. Since the relations for a unity feedback system can be easily extended to a system with a specified feedback transfer function, $H(s)$, the following rules are illustrated by unity feedback systems. For a system with unity feedback from Art. 11-3

$$KG(s) = \frac{T(s)}{1 - T(s)} \quad (11-6-1)$$

where $T(s)$ is the specified over-all closed-loop transfer function. It is also pointed out that the zeros are the same for $KG(s)$ and $T(s)$ so that only the equation

$$1 - T(s) = 0 \quad (11-6-2)$$

is of concern in plotting a locus. In this case the locus to be plotted is one giving all possible roots of Eq. 11-6-2 as the closed-loop gain implicit in $T(s)$ is varied from zero to infinity. Since many of the rules for the direct root locus apply as they stand to the so-called inverse root locus, the rules for plotting this new locus are similar to those used in Art. 11-4.

RULE I-1. REAL AXIS SEGMENTS. From Eq. 11-6-2 it can be seen that the angle criterion to be used for the inverse root locus is that the angle of $T(s)$ be an even multiple of 180° .

$$\text{ang } T(s) = 2\pi j \quad ; \quad j = 0, \pm 1, + 2, \dots \quad (11-6-3)$$

Therefore the real axis segments to the left of an even number of poles and zeros form parts of the inverse root locus.

Example 11-6-1. As a first example involving the inverse root locus

$$T(s) = \frac{K_c(s + 4)}{(s + 10)(s^2 + 10s + 50)} \quad (11-6-4)$$

³ John A. Aseltine, "Feedback System Synthesis by the Inverse Root-Locus Method," *IRE Convention Record*, Part II (1956): 13-17.

⁴ John Zaborsky, "Integrated s -Plane Synthesis Using 2-Way Root Locus," *Trans. AIEE* 76, Part III (January, 1957): 797-801.

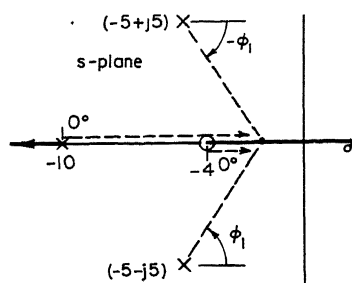


Fig. 11-6-1.

The closed-loop poles and zeros are illustrated in Fig. 11-6-1. Also, the real axis segments that form part of the inverse root locus are shown with heavy lines. For s to the right of -4 the angles for the two factors $s + 4$ and $s + 10$ are both zero while the angles for the two complex-pole terms are equal but opposite. The phase angle of $T(s)$ is zero. For s to the left of -10 the angle of $T(s)$ is 360° , which also satisfies the angle condition above.

RULE I-2. STARTING AND ENDING POINTS. For the inverse root locus

$$|T(s)| = 1 \quad (11-6-5)$$

which is the same as the requirement for the direct root locus. There is no change in Rule 2. Thus, in plotting the inverse root locus, the branches start at *closed-loop* poles and end on *closed-loop* zeros as the closed-loop gain varies from zero to infinity. In Example 11-6-1 branches of the inverse root locus start at $s = -10$, and $-5 \pm j5$ and end on the zero at $s = -4$ and the two zeros at infinity.

RULE I-3. ASYMPTOTES FOR LARGE s . As indicated above, branches of the inverse root locus frequently go to zeros at infinity. The procedure used in Rule 3 for the direct locus can be applied here with the exception that the angles made by the asymptotes become

$$\text{Asymptote angles} = \frac{2\pi k}{n - m} \quad ; \quad k = 0, 1, 2, \dots, n - m - 1 \quad (11-6-6)$$

where n is the highest power of s in the denominator of $T(s)$ and m is the highest power of s in the numerator of $T(s)$.

In establishing a point of origin for the asymptotes, Rule 3 also provides the solution with the point given by

$$\text{Origin of asymptotes} = \frac{\sum_{i=1}^n (\text{Closed-loop poles}) - \sum_{i=1}^m (\text{Finite closed-loop zeros})}{n - m} \quad (11-6-7)$$

Example 11-6-2. Rule I-3 when applied to the system of Example 11-6-1 gives $n - m = 3 - 1 = 2$, indicating two asymptotes that make angles of

$$(2\pi)\left(\frac{0}{2}\right) = 0 \quad \text{and} \quad (2\pi)\left(\frac{1}{2}\right) = \pi \quad (11-6-8)$$

with the positive real axis. These were already arrived at by Rule I-1. No point of origin is needed in this case.

Example 11-6-3. As another example for a unity feedback system, consider

$$T(s) = \frac{K_c}{(s + 10)(s^2 + 6s + 25)} \quad (11-6-9)$$

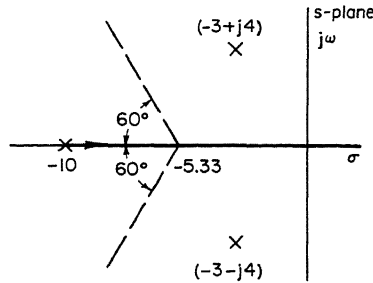


Fig. 11-6-2.

which results in the closed-loop pole-zero configuration of Fig. 11-6-2. There are three zeros at infinity in this case. Therefore, three asymptotes are needed which make angles 0 , $2\pi/3$, and $4\pi/3$ with the positive real axis. Their point of origin is found to be

$$\frac{-10 - 3 + j4 - 3 - j4}{3} = -\frac{16}{3} = -5.33 \quad (11-6-10)$$

RULE I-4. COMPLEX CONJUGATE SEGMENTS. Since normal closed-loop transfer functions, $T(s)$, have real coefficients, their poles and zeros appear in conjugate pairs. Thus, branches of the inverse root locus not on the real s -plane axis occur in conjugate pairs. In Example 11-6-1 there are branches of the inverse root locus that start at $-5 \pm j5$, while in Ex. 11-6-3 branches start at $-3 \pm j4$.

RULE I-5. DEPARTURE OF COMPLEX SEGMENTS FROM THE REAL AXIS. Whenever a segment of the real axis between two like singularities, as between two poles or two zeros, is part of the inverse root locus, two complex branches leave or enter this real segment. The approach that is used to determine breakaway points for the direct root locus can be applied directly to the inverse locus. In the case considered here this means that the change in the angle of $T(s)$ must be zero as s moves from a real axis locus segment to a nearby complex point. The following examples illustrate a method which can be used to find breakaway points for the inverse root locus.

Example 11-6-4. In order to fill in parts of Ex. 11-6-1

$$T(s) = \frac{K_c(s + 4)}{(s + 10)(s^2 + 10s + 50)} \quad (11-6-11)$$

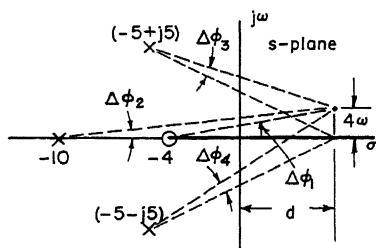


Fig. 11-6-3.

is reconsidered. The partial inverse root locus which is sketched in Fig. 11-6-3 is obtained from previous rules. Since the real axis locus segment from -4 to $+\infty$ is between two zeros, it is necessary for complex branches from the poles at $-5 \pm j5$ to intersect the real axis at some point in the above interval. One branch then goes left to the zero at -4 while the other branch goes right to a zero at infinity. By using the procedure used in previous work in the examples for Rule 5, the following relations can be written.

$$\Delta\varphi_1 - (\Delta\varphi_2 + \Delta\varphi_3 + \Delta\varphi_4) = 0 \quad (11-6-12)$$

$$\frac{1}{d + 4} = \frac{1}{d + 10} + \frac{2(d + 5)}{(d + 5)^2 + 5^2} \quad (11-6-13)$$

$$d = -0.726; -7.64 \pm j3.26 \quad (11-6-14)$$

It can be seen that the intercept is to the left of the origin instead of to the right as assumed in setting up the above equations.

RULE I-6. ANGLES OF DEPARTURE AND ARRIVAL AT COMPLEX POLES AND ZEROS. By now it is almost monotonous to say that the procedure used for the direct root locus can also be applied to the inverse locus. It is fortunate, however, that such is the case in so many of the rules. In the case of Rule I-6 the only change necessary when applying the approach of Rule 6 is that the angle of $T(s)$ must be an even multiple of π radians. Examples 11-6-5 and 11-6-6 illustrate Rule I-6 while filling in information needed for previous examples.

Example 11-6-5. In order to fill in Example 11-6-1, the angle of departure of the inverse locus branch from the pole at $-5 + j5$ is found for

$$T(s) = \frac{K_c(s + 4)}{(s + 10)(s^2 + 10s + 50)} \quad (11-6-15)$$

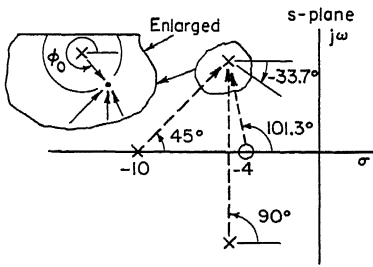


Fig. 11-6-4.

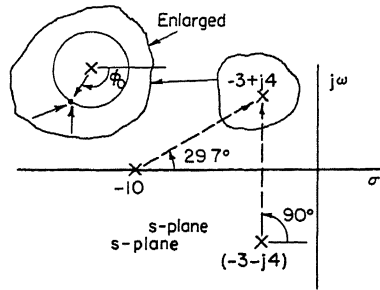


Fig. 11-6-5.

Figure 11-6-4 illustrates the angles used when applying Rule I-6. The angle of $T(s)$ for a point very near the pole at $-5 + j5$ is

$$(101.3^\circ) - (90^\circ + 45^\circ + \varphi_0) = 0^\circ \quad (11-6-16)$$

$$\varphi_0 = -33.7^\circ \quad (11-6-17)$$

where φ_0 is the angle of departure of the locus segment from the pole at $-5 + j5$.

Example 11-6-6. From Example 11-6-3

$$T(s) = \frac{K_c}{(s + 10)(s^2 + 6s + 25)} \quad (11-6-18)$$

The angle of departure of the inverse root locus from the pole at $-3 + j4$ is desired. The angles involved in this case are shown in Fig. 11-6-5. It should be pointed out that the conjugate branch of the locus emanating from the pole at $-3 - j4$ leaves this pole at the negative of the angle found for the branch starting at $-3 + j4$. The relation needed in this case is

$$-(29.7^\circ + 90^\circ + \varphi_0) = 0 \quad (11-6-19)$$

$$\varphi_0 = -119.7^\circ \quad (11-6-20)$$

This angle of departure together with information obtained from the previous examples provides all of the major characteristics needed for making a rough sketch of the complete inverse root locus as shown in Fig. 11-6-6. The complex branches

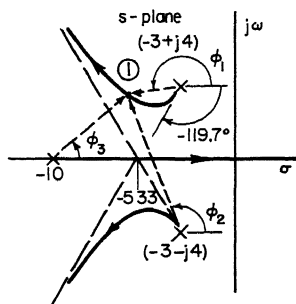


Fig. 11-6-6.

must be checked graphically by measuring angles to points on the rough sketch and adjusting the sketch as necessary to satisfy the angle of $T(s)$ criterion. For example, if point ① in Fig. 11-6-6 is really on the inverse root locus, the angle of $T(s)$ given by $-(\varphi_1 + \varphi_2 + \varphi_3)$ must be zero or an even multiple of 180° . If the point ① on the original rough sketch does not satisfy the above requirement, a new point nearby is chosen and the angle of $T(s)$ redetermined. This process is repeated until enough points which satisfy the angle requirement are found to allow an accurate sketch of the locus.

RULE I-7. CALIBRATION OF THE INVERSE ROOT LOCUS. In addition to the angle condition established by

$$1 - T(s) = 0 \quad (11-6-21)$$

the magnitude condition

$$|T(s)| = 1 \quad (11-6-22)$$

must be satisfied for each specific value of the gain constant, K_c , which is inherent in $T(s)$. As discussed in Rule I-2, the inverse root locus starts at the closed-loop poles for $K_c = 0$ and goes to the closed-loop zeros when $K_c = \infty$. In order to find the value of K_c corresponding to a given point on the accurate sketch of the locus, $T(s)$ is set equal to $K_c T'(s)$ and then

$$K_c = \frac{1}{|T'(s)|} \quad (11-6-23)$$

The magnitude of $T'(s)$ can be determined by measurement using a pair of dividers on the sketch of the locus as in Rule 7. The following example illustrates the method of calibrating the inverse root locus.

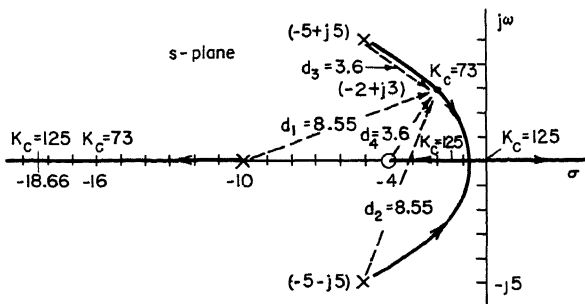


Fig. 11-6-7.

Example 11-6-7. The $T(s)$ used in Ex. 11-6-1 and its inverse root locus which is sketched in Fig. 11-6-7 is used as a first example of the calibration procedure. The complex segments have been determined by using the graphical approach mentioned earlier while employing a Spirule. The point $-2 + j3$ is on the sketch and is chosen as a typical point for which the gain K_c is desired. In other words, what K_c would cause a zero of $1 - T(s)$ to fall at $-2 + j3$? By measuring the

segments d_1 , d_2 , d_3 , and d_4 directly on the diagram the gain K_c is found by Eq. 11-6-23 as

$$K_c = \frac{(8.55)(8.55)(3.6)}{3.6} = 73.1 \quad (11-6-24)$$

From the relation

$$1 - T(s) = s^3 + 20s^2 + (150 - K_c)s + (500 - 4K_c) = 0 \quad (11-6-25)$$

it can be seen that the sum of the roots is -20 for all K_c ; therefore, the location of the third root for $K_c = 73.1$ can be found directly as

$$(-2 + j3) + (-2 - j3) + r_3 = -20 \quad ; \quad r_3 = -16 \quad (11-6-26)$$

This point can be checked easily by using the magnitude relation above since only one length has to be measured — that from $-5 \pm j5$ to -16 . The other lengths are simple numerical differences.

Since a pole of $KG(s)$ at $s = 0$ is frequently desired, the value of K_c that causes such a result is of interest. From Eq. 11-6-25 the necessary conditions is for $500 - 4K_c = 0$ or $K_c = 125$. This means also that the d-c gain of the closed-loop system for this K_c is 1. The other two roots of Eq. 11-6-25 are found from either the quadratic equation remaining after factoring out s or by measurements on the graph and calculation of K_c . These roots turn out to be -18.66 and -1.34 which sum to -20 as required. Generally only a few points need to be calibrated, as in Eq. 11-6-24. This should present no difficulty when using the procedure of Rule I-7.

11-7. Additional Items Concerning the Inverse Root Locus.

While the formulation of the previous rules is based on closed-loop functions with an excess of poles over finite zeros, some mention of other cases should be made. If the number of zeros of $T(s)$ equals or exceeds the number of poles, difficulty is encountered in realizing the open-loop function for negative feedback. Since such cases are not common, this difficulty is mentioned here only with reference to the literature for those interested in further study.⁵

Since the two root loci discussed in this chapter relate open- and closed-loop transfer functions, a further look at this relationship is in order. By combining the two procedures, a method is available for starting with $KG(s)$, first finding $T(s)$ for a given K , and then using this $T(s)$ and going back to the $KG(s)$. Thus, if a given direct root locus for a certain $KG(s)$ gives a certain $T(s)$ for some value of K , the inverse root locus for this $T(s)$ must give back the original $KG(s)$ for a corresponding value of K_c . As mentioned previously, if there are one or more zeros at infinity in $KG(s)$, K and K_c are the same.

Example 11-7-1. If $KG(s)$ for a unity feedback control system is

$$KG(s) = \frac{K}{s(s+4)} \quad (11-7-1)$$

⁵ John A. Aseltine, *loc. cit.*

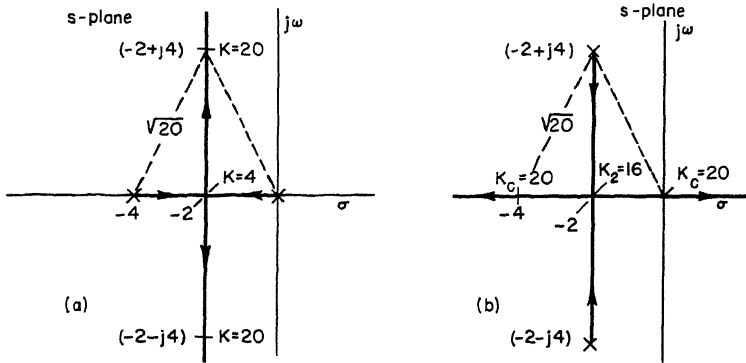


Fig. 11-7-1.

the direct root locus of Fig. 11-7-1a results. For a system gain constant $K = 20$ the closed-loop poles are at $-2 \pm j4$ as shown. The $T(s)$ function can then be written as

$$T(s) = \frac{20}{s^2 + 4s + 20} \quad (11-7-2)$$

However for $T(s)$ specified as

$$T(s) = \frac{K_c}{s^2 + 4s + 20} \quad (11-7-3)$$

the inverse root locus is shown in Fig. 11-7-1b. If K_c is now set at 20, open-loop poles result at -4 and 0 as indicated giving a

$$KG(s) = \frac{20}{s(s + 4)} \quad (11-7-4)$$

In this case K and K_c are the same. Also, the roots move along the two loci in *opposite* directions as K and K_c increase. While the two loci include common parts between 0 and -4 and $-2 \pm j4$, they also include additional sections. The direct root locus includes *all* possible values of s for closed-loop poles which result from a given open-loop function as K varies. The inverse or reverse root locus shows all possible open-loop poles that result from a given set of closed-loop poles as K_c varies.

It should now be apparent that a circular process results from combining the direct and inverse root loci. This will be an advantage later in system design.

11-8. The Gain Adjustment Problem. While much has been said about constructing the direct and inverse root loci, nothing has been presented regarding the application of these tools in analysis and design. Generally the first step in such application work is to interpret the system specifications in terms of closed-loop poles and zeros. The design then proceeds by trying to reconcile the characteristics of certain major fixed system components with the desired closed-loop poles and zeros. The

simplest case arises when a gain adjustment in the open-loop system gives suitable closed-loop poles and zeros. No alteration of the basic system dynamics is needed in this case. A straightforward use of the direct root locus makes evident whether or not a desired set of closed-loop poles and zeros can be obtained by gain adjustment.

Example 11-8-1. For a system of the form shown in Fig. 11-8-1 it is desired to determine whether or not the following over-all performance specifications can be met by gain adjustment only: $K_v = 6 \text{ sec}^{-1}$, overshoot for unit-step input ≤ 25 per cent, rise time $\leq 0.3 \text{ sec}$.

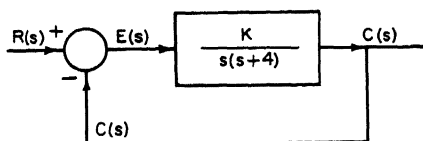


Fig. 11-8-1.

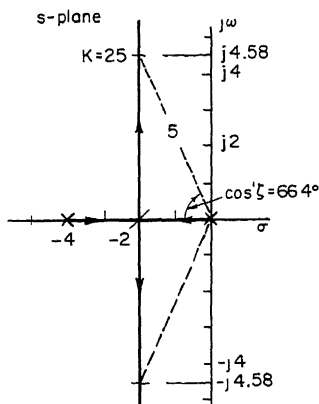


Fig. 11-8-2.

Since the system is of second order, the curves of Chapter 6, Fig. 6-11-3 can be used to relate overshoot, etc., to damping ratio, ζ . A $\zeta \approx 0.4$ is found to provide suitable overshoot and so a line making an angle of $\psi = \cos^{-1} \zeta$ with the negative real axis is drawn on an s -plane plot containing the direct root locus for Fig. 11-8-1. From the direct root locus of Fig. 11-8-2 it can be seen that the given system results in closed-loop poles at $-2 \pm j4.58$ for a damping ratio of 0.4 with an undamped natural frequency, ω_n , of 5 radians/sec. From the curves of Fig. 6-11-2, the normalized 10-90 per cent rise time, $\omega_n T_R$, is seen to be approximately 1.5 radians. Unnormalizing this gives a rise time of $T_R = 0.3$ seconds. Finally the velocity constant can be found as $K_v = K/4$ which, for a $\zeta = 0.4$, is $25/4 = 6.25$. Therefore, the three specifications are all met merely by setting the system gain at $K = 25$ with no modifications to the open-loop pole-zero pattern. This simple case could have been handled directly without the direct root locus, but it provides valuable knowledge of closed-loop pole position vs. gain. For more involved systems the advantages of the direct root locus are more apparent since many more closed-loop poles can be observed simultaneously.

11-9. Lead Compensation by Using the Direct-Root-Locus Approach. For some cases in which the fixed part of the open-loop pole-zero pattern does not lead to satisfactory closed-loop poles and zeros a simple lead compensation network can be inserted in series in the forward gain path. The goal in this scheme is to provide leading phase shift in the open-loop transfer function in the region of frequencies that normally result in 180° lagging phase angle. This is depicted in Fig. 11-9-1 which shows part of a simple Nyquist diagram. When this

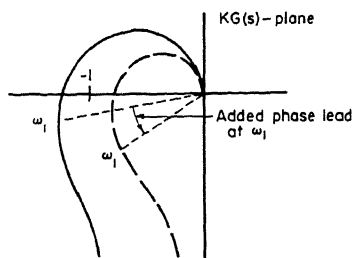


Fig. 11-9-1.

technique is considered earlier in Chapters 8 and 9, it is approached from the Nyquist and Bode diagram standpoints. The lead effect on the open-loop phase is quite apparent in such schemes. The effect of a simple lead network, whether it be electrical, mechanical, or pneumatic, can be seen in terms of the direct root locus by adding a zero and a pole along the negative real axis with the zero nearer the origin. The fact that this scheme gives leading phase shift can be demonstrated by letting s travel up the $j\omega$ -axis while observing the phase angle of the compensation function

$$G_1(s) = K_1 \frac{s + z}{s + p} \quad (11-9-1)$$

In order to see the effect on the root locus, the following example is considered.

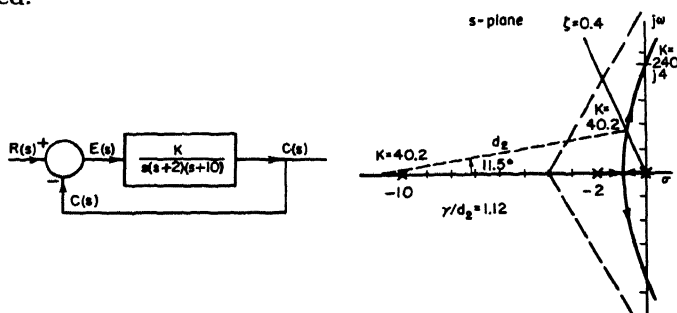


Fig. 11-9-2.

Example 11-9-1. The system of Fig. 11-9-2a is to be modified by use of a lead network. It is desired to obtain a faster rise time without severe overshoot, together with a larger velocity constant. Normally the specifications would include phase margin restrictions when the work is carried out on the Nyquist diagram or the Bode plot. However, an attempt to show only the general effect of a simple lead network on the root locus is undertaken here. In order to fix values, an overshoot in the order of 25 per cent is allowable. From the transient step response curves of Chapter 6 for a second-order system it is found that a $\zeta = 0.4$ would meet this requirement. Since Fig. 11-9-2a is a third-order system, this is only approximately valid; however, generally a third-order pole stabilizes an oscillatory response.

Digressing for a moment, the following general Type I third-order closed-loop transfer function is considered for a step input:

$$C(s) = \frac{\gamma \omega_n^2}{s(s + \gamma)(s^2 + 2\zeta \omega_n s + \omega_n^2)} \quad (11-9-2)$$

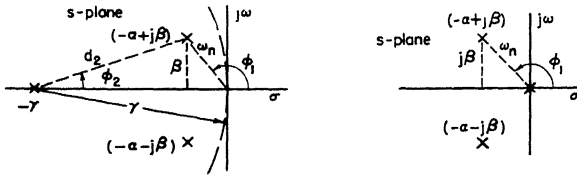


Fig. 11-9-3.

Some appropriate geometrical terms are shown in Fig. 11-9-3a and b in order to simplify the inversion of $C(s)$ to $c(t)$. For the third-order system

$$c(t) = 1 - \frac{\omega_n^2}{d_2^2} e^{-\gamma t} + \frac{\gamma \omega_n}{d_2 \beta} e^{-\alpha t} \sin(\beta t + \psi_1) \quad (11-9-3)$$

with $\psi_1 = -\varphi_1 - \varphi_2$. For a second-order system with the same complex pole pair (same ζ and ω_n),

$$c(t) = 1 + \frac{\omega_n}{\beta} e^{-\alpha t} \sin(\beta t - \varphi_1) \quad (11-9-4)$$

The amplitude and phase of the oscillatory terms of Eqs. 11-9-3 and 11-9-4 can be related as follows:

$$\frac{\text{Third order}}{\text{Second order}} = \frac{\gamma}{d_2} \angle \varphi_2 \quad (11-9-5)$$

Thus if $\gamma < d_2$ the oscillatory term is certainly reduced, but $-\varphi_2$ is also important since it shifts the maximum value of the sinusoid as indicated in Fig. 11-9-4. While

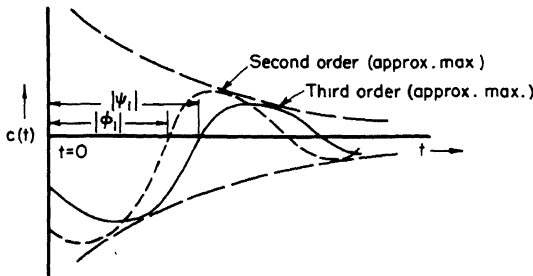


Fig. 11-9-4.

the scale change γ/d_2 is not shown in this figure, it does point out that a reduction in overshoot can result even though $\gamma > d_2$. Actually the circular arc of radius γ centered at $-\gamma$ as drawn in Fig. 11-9-3a quickly shows whether γ is greater or less than d_2 . Furthermore, the exponential term due to the pole at $-\gamma$ subtracts from the oscillatory term. If γ is much greater than α , this exponential term decays rapidly and contributes little to the overshoot.

Returning to the original example, a line for $\zeta = 0.4$ is drawn on the root-locus diagram of Fig. 11-9-2b and the intersections with the root-locus complex branches are found. Thus the closed-loop transfer function would have to be

$$\frac{C(s)}{R(s)} = \frac{40.2}{(s + 10.4)[(s + 0.78)^2 + (1.8)^2]} \quad (11-9-6)$$

The remote location of $-\gamma$ means that it has little effect, and the response $c(t)$ for a unit step gives a rise time using second-order curves of about 0.75 seconds, an overshoot of about 25 per cent and K_v is found as approximately 2.

If a simple lead network is added, one goal is to speed up the response as generally associated with raising the system bandwidth. At the same time it is desirable to hold the overshoot figure and raise the velocity constant. Normally, the lead network is set to give maximum phase lead slightly above the frequency at which the uncompensated system open-loop function has 180° phase lag. The added phase lead must be sufficient to give the necessary phase margin. In the root-locus method only the frequency at which the lag is 180° is readily apparent as evidenced by the frequency at which the root locus crosses the imaginary s -plane axis. This aids in locating the lead zero and pole. It is easy to show that maximum lead for a compensation function of the form of Eq. 11-9-1 occurs at the geometric mean of the two break frequencies. If the lead function is written as

$$G_1(s) = K \frac{s + a}{s + ka} \quad ; \quad k > 1 \quad (11-9-7)$$

the maximum lead occurs at $\omega_m = a\sqrt{k}$ and the maximum phase shift is $\tan^{-1}[(k - 1)/2\sqrt{k}]$. A convenient k is 10 giving a maximum lead of 54.9° . Since the root locus of the original system crosses the imaginary axis at $\omega = 4.48$, a rough guess for the parameter a is $a \geq 4.48/\sqrt{10}$. Due to the increase in phase lag with frequency a is set equal to 2 to cancel the pole at -2 . This leaves a compensated function

$$KG(s) = \frac{K}{s(s + 10)(s + 20)} \quad (11-9-8)$$

While the new root locus can be easily constructed, a look at the lead network pole and zero in the s -plane reveals the main effect on the original root locus. In Fig. 11-9-5a the angle of $G_1(s)$ of Eq. 11-9-7 is constructed for a general complex value of $s = s_1$ and the factors $s_1 + a$ and $s_1 + ka$ are shown. The angle of $G_1(s)$ of Eq. 11-9-7 is seen to be a leading angle $\phi_1 - \phi_2$. The effect of $G_1(s)$ on the root locus of a

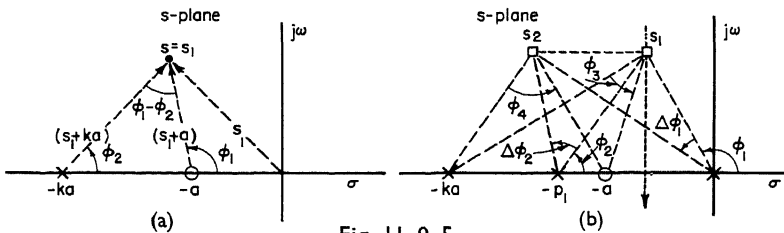


Fig. 11-9-5.

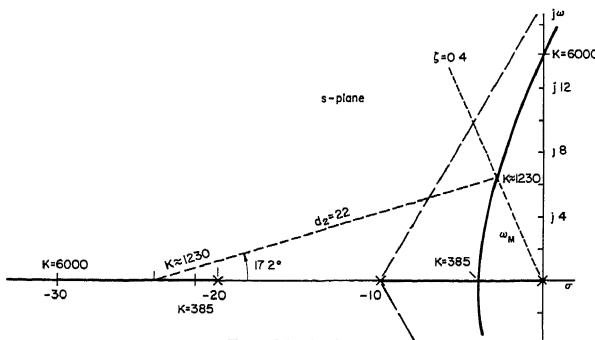


Fig. 11-9-6.

function with poles at the origin and at $-p_1$ is shown in Fig. 11-9-5b. It can be seen that a given point on the original root locus such as s_1 must be shifted to the left to s_2 , so that the lead angle is offset by the increase in the lag angles φ_1 and φ_2 . The direct root locus for Eq. 11-9-8 results in Fig. 11-9-6, which also shows the line for a constant damping ratio $\zeta = 0.4$. If the gain is set at approximately $K = 1230$, the closed-loop poles are at $-3 \pm j6.5$ and -24 , so that

$$\frac{C(s)}{R(s)} = \frac{1230}{(s + 24)[(s + 3)^2 + (6.5)^2]} \quad (11-9-9)$$

A comparison of Figs. 11-9-6 and 11-9-2b makes it apparent that the root locus has been shifted to the left, as discussed above. From Eq. 11-9-5 the oscillatory term in the step response for the system of Eq. 11-7-9 is altered from that for a second-order system with the same complex poles by the factor

$$\frac{\gamma}{d_2} \angle -\varphi_2 = 1.09 \angle -17.2^\circ \quad (11-9-10)$$

The phase lag causes the maximum to shift approximately

$$\Delta t = \frac{17.2^\circ \times \pi}{180^\circ \times \beta} = 0.0462 \text{ sec} \quad (11-9-11)$$

with the exponential $e^{-\alpha t}$ decreasing by a factor

$$e^{-(\Delta t)(\beta)} = e^{-0.139} = 0.87 \quad (11-9-12)$$

in this interval. Since the maximum overshoot in the second-order system occurs at approximately $\omega_n t = 3.4$, from the curves of Fig. 6-11-2, $t_{\max}$ should now be at approximately

$$\frac{3.4}{7.1} + 0.0462 \approx 0.52 \text{ sec} \quad (11-9-13)$$

and the term

$$\frac{\omega_n^2}{d_2^2} e^{-\gamma t} = \frac{(7.1)^2}{(22)^2} e^{-(24)(0.52)} \quad (11-9-14)$$

from Eq. 11-9-4 is negligible. Thus the overshoot should be about $(1.09)(.87) \approx 0.95$ times that for a second-order system with $\zeta = 0.4$. Therefore, 25 per cent is still a good overshoot figure.

The velocity constant is now

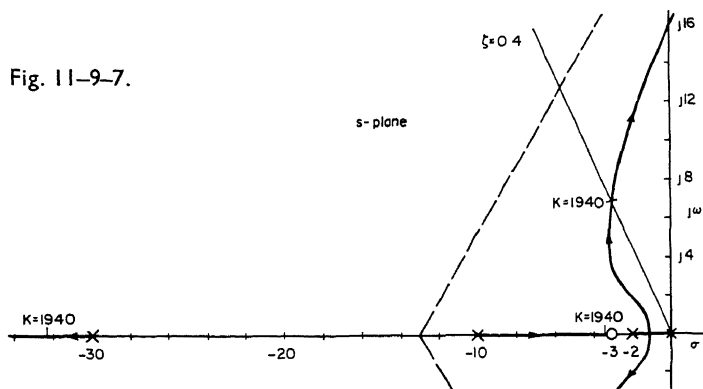
$$K_v = \frac{1230}{200} = 6.15 \quad (11-9-15)$$

By neglecting the effect of the pole at -24 , the rise time is found from second-order system curves to be about 0.2 seconds.

All of the effects of the lead network on system response are due to moving the closed-loop poles away from the origin. The bandwidth has been increased, so that higher signal (and noise) frequencies can be transmitted through the system.

If the lead pole and zero had been moved out to say -30 and -3 respectively, the alteration in the root locus is more apparent while there is little change in the response. Fig. 11-9-7 shows the results. The response term due to the pole at approximately -3.4 , while it decays slowly, is of very small amplitude compared with the oscillatory response term.

Fig. 11-9-7.



11-10. Lag Compensation and the Direct Root Locus. As discussed in Chapters 8 and 9 lag or integral compensation is mainly used to allow an increase in the low-frequency gain without altering the high-frequency behavior. The added pole and zero must not contribute appreciable phase lag at frequencies near cutoff for the system. The only way to accomplish this is to locate the compensating pole and zero very close to the origin as compared with other system poles and zeros, with the added pole inside the zero. In this case the compensation function can be written

$$G_1(s) = K \frac{s + ka}{s + a} ; \quad k > 1 \quad (11-10-1)$$

and the frequency of maximum phase shift is still $a\sqrt{k}$. The maximum phase shift at this frequency is $-\tan^{-1}[(k-1)/2\sqrt{k}]$. In order to see the effect on the direct root locus, Example 11-10-1 is considered.

Example 11-10-1. The open-loop function which is used in Example 11-9-1, namely

$$KG_2(s) = \frac{K}{s(s+2)(s+10)} \quad (11-10-2)$$

is reconsidered. It is desired to increase the velocity constant by 10 times while not appreciably altering an overshoot of about 25 per cent. The compensated function has a velocity constant given by

$$K_v = \frac{K(k)}{20} \quad (11-10-3)$$

where k is from Eq. 11-10-1. Thus $k = 10$ would appear suitable. To prevent altering the root locus near the original system closed-loop poles and zeros (those for $\zeta = 0.4$ in Fig. 11-9-2b, the added pole and zero have to appear to merge from the complex pair of poles at $-0.78 \pm j1.8$. While the exact location of the compensation pole and zero is not critical, there is no need to use excessively small values, running into component trouble. Therefore, the zero is located at -0.1 and the pole at -0.01 . From the complex closed-loop poles of Fig. 11-9-2b this gives an

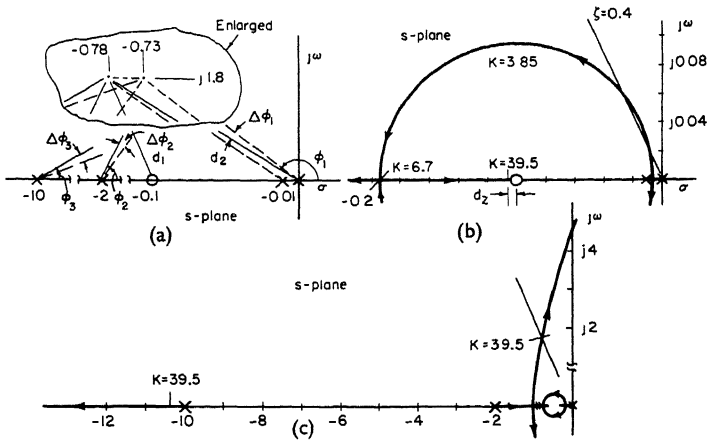


Fig. 11-10-1.

additional lag of less than 3° as shown in Fig. 11-10-1a. This means that the root locus at this point for the compensated system is shifted to the right decreasing the lag (pole) angles just enough to offset this compensation lag. The gain K at $-0.78 \pm j1.8$ is not changed appreciably since $d_1 \approx d_2$ and these are the only really new factors in determining K . Near the origin there is considerable change as shown in Fig. 11-10-1b which pictures an enlarged version of the region of the s -plane around the origin. The poles at -2 and -10 contribute negligible phase angles in this region and are neglected in everything except the determination of K . It can be seen from Fig. 11-10-1b that a closed-loop pole lies in the vicinity of the origin between -0.1 and -0.195 . However, the exponential response term that results from this pole, while it decays slowly, is small because of the nearby zero at -0.1 (see length d_z in Fig. 11-10-1b).

The over-all compensated root locus is shown in Fig. 11-10-1c with the closed-loop transfer function for a $\zeta = 0.4$ given by

$$\frac{C(s)}{R(s)} = \frac{39.5(s + 0.1)}{(s + 0.1047)(s + 10.41)[(s + 0.75)^2 + 1.75^2]} \quad (11-10-4)$$

The pole at -0.1047 is found by trial and error, using a known gain of 39.8. In fact neither exponential term due to the two real poles contributes much to the time response. The final open-loop function is

$$KG(s) = \frac{(39.5)(s + 0.1)}{s(s + 0.01)(s + 2)(s + 10)} \quad (11-10-5)$$

from which a velocity constant of 19.8 is realized, which is approximately that desired. It might be mentioned that in cases like this a slightly larger lag-network time-constant ratio would have given the desired factor-of-10 improvement in the velocity constant.

11-11. General Compensation Using the Direct Root Locus.

While the direct root locus makes the open-loop to closed-loop pole relationships directly available for a given set of open-loop singularities, there is still considerable work involved in carrying out a reasonably exact correlation of the closed-loop time response with the closed-loop pole-zero pattern. While general relations are easily established, more exact de-

termination of such things as rise time, settling time and overshoot for all but simple cases involves determining the time-response equation and further plotting or calculation from this point. Frequently several terms can be thrown out as causing negligible effect on the time response. By checking ratios of the lengths which represent various factors in the coefficients of the time-response terms, those that are negligible can be determined.

In order to increase the rise time and settling time, it is generally necessary to move the closed-loop poles to the left by some means such as canceling and unwanted open-loop pole with a compensation zero and then introducing a new pole farther to the left. At the same time, this generally shifts the direct root locus and the closed-loop poles. Movement of the closed-loop poles to the left also increases the velocity constant as can be seen by studying the relation between the closed-loop poles γ_i , zeros $-\sigma_i$, and K_v , which is given by

$$\frac{1}{K_v} = \sum_i \frac{1}{\gamma_i} - \sum_i \frac{1}{\sigma_i} \quad (11-11-1)$$

If all closed-loop poles and zeros are multiplied by 10, K_v is increased 10 times. Another way to increase K_v in some cases, without shifting closed-loop poles and zeros, is to use the lag-network scheme of Art. 11-10. The student should analyze the effect of the lag network on Eq. 11-11-1.

In compensating systems, a check on the number of added poles and zeros must be maintained so that the final compensated open-loop function and the uncompensated open-loop function have the same or a greater excess of finite poles over finite zeros. This prevents the compensation function from having a pole at infinity, which is physically unobtainable.

11-12. Use of the Inverse Root Locus. Ideally it would be desirable in designing control systems to specify certain overall requirements, from these requirements derive a set of closed-loop poles and zeros, and from these poles and zeros obtain the open-loop poles and zeros needed. However, there are several problems not readily solved in this scheme. First, interpreting the commonly used specifications in terms of a set of closed-loop poles and zeros is not a unique process nor is it easily done in most cases. In other words, more than one set of closed-loop poles will meet a given set of specifications. For certain configurations of closed-loop poles and zeros considerable study has resulted in normalized transient response curves for step inputs which allow choosing suitable over-all transfer functions. These cases are primarily the cases involving

1. Two closed-loop poles (Chapter 6)
2. Two closed-loop poles and one zero⁶
3. Three closed-loop poles⁶

4. Three closed-loop poles and one zero^{6, 7}
5. Three closed-loop poles and two complex zeros⁶
6. Four complex closed-loop poles⁶

The addition of a finite zero to the basic second-order system serves to increase the velocity constant, bandwidth, and overshoot. In many cases these effects on bandwidth and overshoot are objectionable. The addition of a third closed-loop pole on the negative real axis is useful in counteracting these effects while also allowing more complicated open-loop systems to be included. Addition of many more poles and zeros soon leads to more confusion than can be easily manipulated. Therefore, unless the fixed part of the open-loop system has more than three poles and one zero the configuration of Case 4 above is generally a suitable choice for a closed-loop transfer function.

Assuming that a set of closed-loop poles and zeros can be found, the next problem is to determine what open-loop blocks are needed. This is where the inverse root locus enters the picture since it presents graphically the desired information. If a suitable choice of open-loop poles on the inverse root locus cannot be made, because of some existing components, a cyclical process of adjustments of both open- and closed loop singularities is now possible, using both direct and inverse root loci as guides. The symmetry in this procedure is pointed out by Zaborsky.⁸ This point is best illustrated by considering an example.

Example 11-12-1. As a problem illustrating the above discussion a unity-feedback-control system is to be designed to meet the following specifications:

Velocity constant ≥ 50

Overshoot for unit-step input ≤ 20 per cent

Rise time ≤ 0.1 sec.

BW ≤ 30 radians/sec.

A fixed part of the system with a transfer function

$$KG_2(s) = \frac{K_1}{s(s + 10)} \quad (11-12-1)$$

is assumed. Investigating $KG_2(s)$ only with the loop closed for a maximum overshoot of 20 per cent gives a $K_1 = 123$, a velocity constant of 12.3, a rise time of 0.15 seconds, and a BW of 14.7 radians/second. Apparently the velocity constant as well as the rise time must be improved. While a simple added zero might suffice,

⁶ Olle I. Elgerd and William C. Stephens, "Effect of Closed-Loop Transfer Function Pole and Zero Locations on the Transient Response of Linear Control Systems," *Trans. AIEE* 77 (May, 1959): 121-127.

⁷ C. R. Hausenbauer and G. V. Lago, "Synthesis of Control Systems Based on Approximation to a Third Order System," *Trans. AIEE* 77 (November, 1958): 415-421.

⁸ John Zaborsky, *loc. cit.*

it would lead to difficulties in realizing the compensation device and so the closed-loop system function is chosen with a dominant pair of complex poles, a negative real pole, and a finite negative real zero. To start with the over-shoot problem is considered. Since the added zero increases the overshoot from the values given by a dominant pole pair alone, an overshoot considerably below the 20 per cent figure allowed is used to help locate the dominant poles. From Fig. 6-11-3 on overshoot for second-order systems, it is seen that $\zeta = 0.707$ gives about 5 per cent overshoot and this is used as a starting point. Additionally, from simple second-order system equations

$$T_{\text{peak}} = \frac{\pi}{\omega_n \sqrt{1 - \zeta^2}} = \frac{\pi}{\omega_{\text{damped}}} \quad (11-12-2)$$

Since the rise time is of the order of $\frac{1}{2}T_p$, $\omega_{\text{damped}} \approx 16$ is chosen here. This means that the complex poles are located at $-16 \pm j16$. For the added zero and pole the velocity constant equation gives

$$\frac{1}{K_v} = \frac{1}{\gamma_1} + \frac{1}{\gamma_2} + \frac{1}{\gamma_3} - \frac{1}{\sigma_1} \quad (11-12-3)$$

$$\frac{1}{50} = \frac{1}{16} + \frac{1}{\gamma_3} - \frac{1}{\sigma_1} \quad (11-12-4)$$

The pole γ_3 should be far enough out that it does not appreciably affect the transient response, whereas the zero cannot be too far in. For simplicity $\sigma_1 = 16$ and $\gamma_3 = 50$ are chosen so that a rough closed-loop transfer function is

$$T(s) = \frac{K_c(s + 16)}{(s + 50)[(s + 16)^2 + (16)^2]} \quad (11-12-5)$$

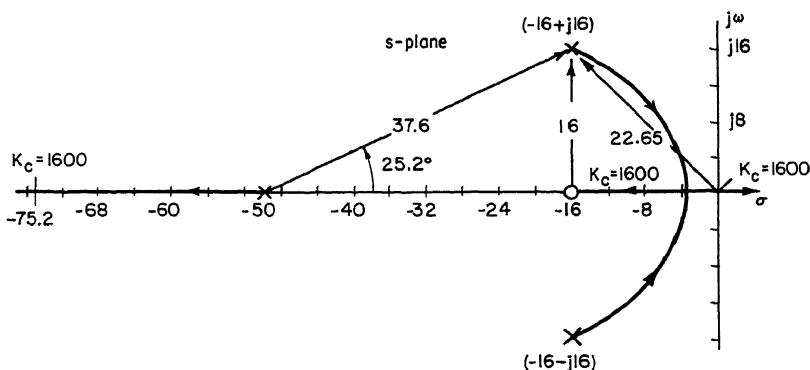


Fig. 11-12-1.

It is now time to sketch the inverse root locus as shown in Fig. 11-12-1. For an open-loop pole at $s = 0$, $K_c = 1600$ and the open-loop function is

$$KG(s) = \frac{1600(s + 16)}{s(s + 6.8)(s + 75.2)} \quad (11-12-6)$$

While this transfer function could be used, it would require that the compensation device cancel the pole at -10 as well as add the zero at -16 and the pole at -75.2 . Actually it may be possible to shift the pole at -6.8 to -10 without too adverse an effect on the closed-loop pole-zero pattern. Therefore such a shift is made as well

as a shift in the pole at -75.2 to -100 in order to move the closed-loop pole further out. A direct root locus of the new function is now made, as shown in Fig. 11-12-2:

$$KG(s) = \frac{K(s + 16)}{s(s + 10)(s + 100)} \quad (11-12-7)$$

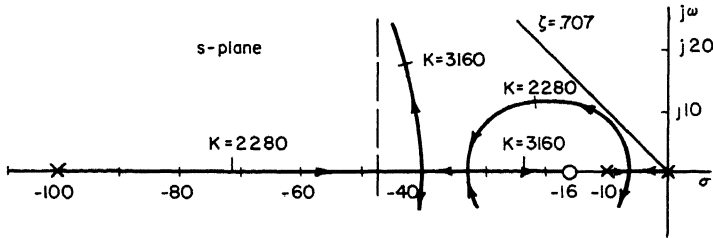


Fig. 11-12-2.

For a velocity constant of 50, $K = 3125$, which throws the complex poles out beyond -40 and brings a real pole in near -25 . The bandwidth is also excessive now, as can be found from a quick Bode plot of $T(j\omega)$ for a K near 3100. The unit-

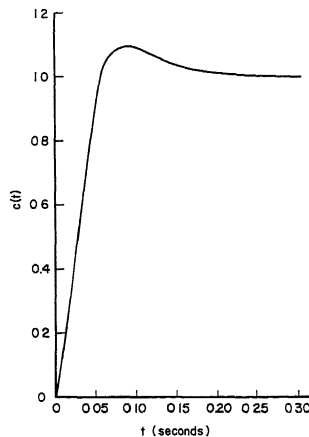


Fig. 11-12-3.

step response for $K = 3175$ is shown in Fig. 11-12-3. The BW is 53 radians/sec for this case. It can be seen from Fig. 11-12-2 that if the zero at -16 had been moved out, the root locus would have been more suitable in shape. However the closed-loop poles are adjusted again, this time keeping the original pole at -50 and using complex poles at $-20 \pm j20$ and a zero at -20 . An inverse root locus of this configuration is drawn as shown in Fig. 11-12-4.

$$T(s) = \frac{K_c(s + 20)}{(s + 50)[(s + 20)^2 + (20)^2]} \quad (11-12-8)$$

For an open-loop pole at $s = 0$, $K_c = 2000$ and

$$KG(s) = \frac{(2000)(s + 20)}{(s)(s + 10)(s + 80)} \quad (11-12-9)$$

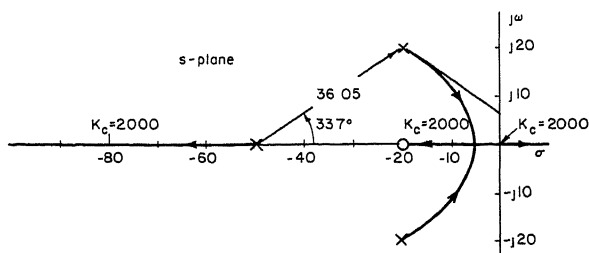


Fig. 11-12-4.

which includes the pole at -10 . The BW of 42 is slightly excessive but this point is assumed not to be of major importance here. The required compensation is

$$KG_1(s) = \frac{KG(s)}{G_2(s)} = \frac{2000(s+20)}{K(s+80)} \quad (11-12-10)$$

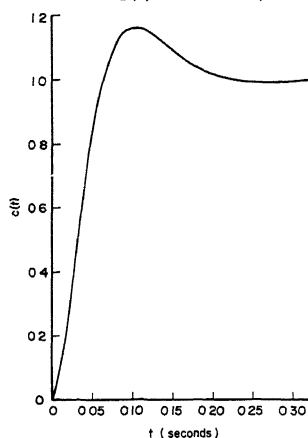


Fig. 11-12-5.

The unit-step response for this system is given in Fig. 11-12-5, which shown the overshoot to be about 16 per cent and the rise time to be 0.046 second.

If the zero had been moved to -20 with the open-loop pole kept at -100 , the direct root locus

$$KG(s) = \frac{K(s+20)}{s(s+10)(s+100)} \quad (11-12-11)$$

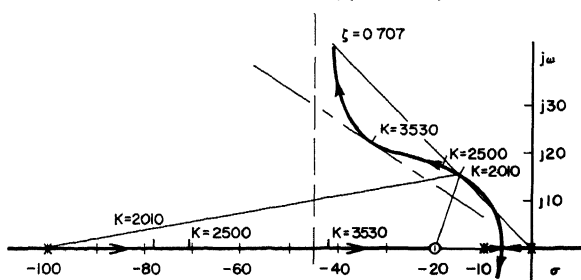


Fig. 11-12-6.

would be as shown in Fig. 11-12-6. For a velocity constant of 50, $K = 2500$ and the closed-loop poles fall at -70.3 , $-19.9 \pm j17.8$. The step responses for this gain setting and for $K = 3530$ are shown in Fig. 11-12-7. A curve for $K = 2010$ is also shown in this figure. The bandwidth for $K = 2500$ is about 44.

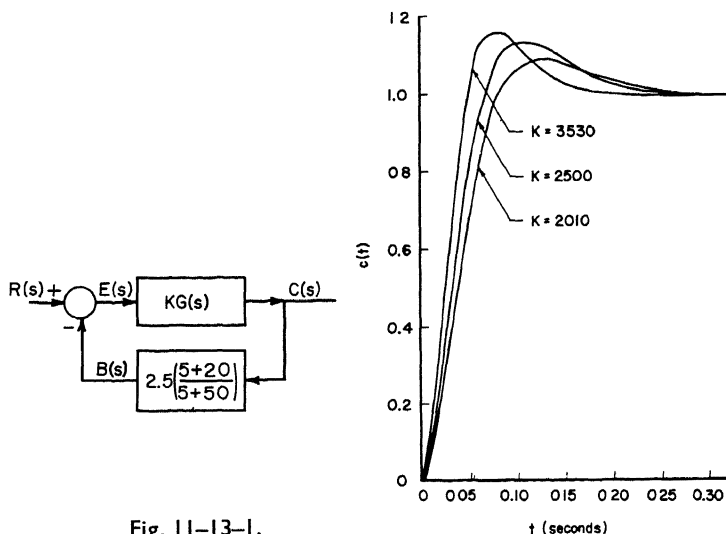


Fig. 11-13-1.

Fig. 11-12-7.

While the above example illustrates the cyclical procedure for using the inverse and direct root loci, it does not necessarily give the most desirable solution. With experience, an ability to correlate pole-zero patterns with various responses and loci patterns will develop making it easier to carry out the required design. Extreme accuracy is not necessary in making the locus sketches required to refine the design. Only the final sketch (which is generally a direct root locus) needs refinement.

11-13. Non-Unity Feedback Systems. So far very little has been said about applying the direct and inverse root loci to control systems which have feedback elements $H(s)$. The direct root locus procedure must be applied to $KG(s)H(s)$, while the inverse root locus must be applied to $T(s)H(s)$ as pointed out in Art. 11-3. The poles of $H(s)$ appear as zeros of $T(s)$ when using the direct root locus. Similarly the poles of $H(s)$ appear as zeros of $KG(s)$ in applying the inverse root locus. Example 11-13-1 illustrates the inverse root locus procedure as applied to a non-unity feedback system.

Example 11-13-1. Fig. 11-13-1 is used as an example of a closed-loop transfer function with feedback elements.

Use of the inverse root locus to find $KG(s)$ means that the procedure for this locus must be applied to $T(s)H(s)$ as indicated above.

$$T(s)H(s) = \frac{2.5K_c(s + 20)}{(s + 33.4)[(s + 13.3)^2 + (20)^2]} \quad (11-13-1)$$

The result of this procedure is shown in Fig. 11-13-2 from which $K_c = 386$ for an

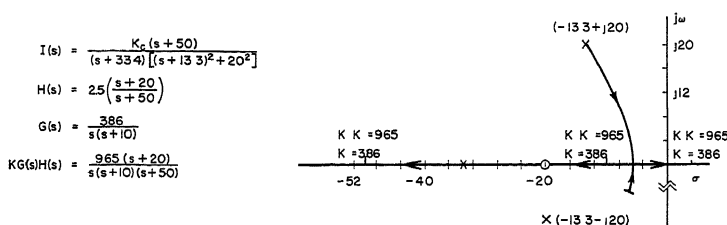


Fig. 11-13-2.

open-loop pole at the origin. The other poles are at -10 and -50 . $KG(s)H(s)$ has the same zeros as $T(s)H(s)$. Therefore

$$KG(s)H(s) = \frac{(2.5)(386)(s+20)}{s(s+10)(s+50)} \quad (11-13-2)$$

and

$$KG(s) = \frac{386}{s(s+10)} \quad (11-13-3)$$

The appearance of the various gain constants should be checked in this example.

11-14. Concluding Remarks. The two root-locus methods presented in this chapter are only the simple forms of a general approach to the study of system equations. Any system parameter can be varied in studying root loci in more general problems. However, these problems are beyond the scope of the presentation here and the reader is referred to more advanced texts or to the literature. Also only ratios of polynomials have been considered. Inclusion of dead-time delays⁹ e^{-sT} and distributed time delays $e^{-\sqrt{s}T}$ in system transfer function can also be handled by the root-locus method. In fact, such systems are probably best analyzed this way. Here again, this material is left for more advanced work.

PROBLEMS

11-1. Plot the roots of the following equations in the s -plane as K varies from 0 to ∞ .

- | | |
|-----------------------------|---|
| (a) $1 + \frac{K}{s} = 0$ | (e) $\frac{K}{s(s+1)} = -1$ |
| (b) $1 + \frac{K}{s^2} = 0$ | (f) $\frac{K(s+1)}{(s+2)(s+3)} + 1 = 0$ |
| (c) $1 + \frac{K}{s^3} = 0$ | (g) $Ke^{-s} + 1 = 0$ |
| (d) $\frac{K}{s+1} = -1$ | (h) $1 - \frac{K}{(s+10)(s+3)} = 0$ |

⁹ Yaohan Chu, "Feedback Control Systems with Dead-Time or Distributed Lag by Root-Locus Method," *Trans. AIEE* 71, (November, 1952): 291-296.

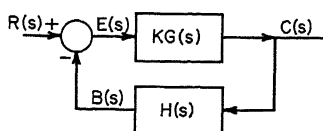


Fig. 11-P-2.

11-2. Determine the relationships between the open- and closed-loop poles and zeros for the system of Fig. 11-P-2. Don't solve any equations for roots, but set up all equations that are needed. Note how open-loop gains, zeros, and poles appear.

$$(a) \quad KG(s) = \frac{100}{s(s+10)} \quad ; \quad H(s) = 1$$

$$(b) \quad KG(s) = \frac{K(s+10)}{s^2(s+20)} \quad ; \quad H(s) = \frac{K_h s}{(s+10)^2}$$

$$(c) \quad KG(s) = \frac{Ke^{-2s}}{s} \quad ; \quad H(s) = K_h s$$

11-3. Construct the direct root locus for each of the unity feedback systems with the following $KG(s)$ transfer functions. Make use of the appropriate Rules 1 through 7 in this chapter.

$$(a) \quad KG(s) = \frac{K(s+1)(s+3)}{s(s+2)}$$

$$(b) \quad KG(s) = \frac{K}{s[(s+5)^2 + (12)^2]}$$

$$(c) \quad KG(s) = \frac{K(s+2)}{s^2(s+3)}$$

$$(d) \quad KG(s) = \frac{K}{s(s+10)[(s+0.01)^2 + (0.01)^2]}$$

$$(e) \quad KG(s) = \frac{K}{(s+1)(s+4)(s+100)}$$

$$(f) \quad KG(s) = \frac{K(s+0.05)}{s(s+0.4)(s+10)}$$

$$(g) \quad KG(s) = \frac{K}{(s+100)^2(s+150)}$$

$$(h) \quad KG(s) = \frac{K[(s+1)^2 + 1^2]}{s(s+1)(s+4)}$$

$$(i) \quad KG(s) = \frac{K(s-4)}{s(s+4)}$$

$$(j) \quad KG(s) = \frac{Ke^{-s}}{s}$$

11-4. Construct the direct root locus for each of the following systems. Assume a basic negative feedback connection.

$$(a) \quad KG(s) = \frac{K}{s(s+10)} \quad ; \quad H(s) = K_h s$$

$$(b) \quad KG(s) = \frac{K}{s^2(s+4)} \quad ; \quad H(s) = \frac{K_h s}{(s+1)(s+2)}$$

$$(c) \quad KG(s) = \frac{K(s+20)}{s^2(s+30)} \quad ; \quad H(s) = \frac{K_s}{(s+40)(s+50)}$$

11-5. Construct the inverse root locus for each of the following unity negative feedback control systems.

$$(a) \quad T(s) = \frac{K_c}{s^2 + 4s + 25}$$

$$(b) \quad T(s) = \frac{K_c(s+10)}{(s+15)(s^2+2s+2)}$$

$$(c) \quad T(s) = \frac{K_c(s^2 + 377^2)}{(s^2 + 320s + 400^2)(s+500)}$$

$$(d) \quad T(s) = \frac{K_c(s+100)}{s^2 + 20s + 625}$$

$$(e) \quad T(s) = \frac{K_c}{(s+10^5)(s+5 \times 10^4)(s+4 \times 10^4)}$$

$$(f) \quad T(s) = \frac{K_c(s+0.1)}{(s+0.2)(s+100)(s^2+40s+400)}$$

11-6. Construct the inverse root locus for the following non-unity feedback systems and determine $KG(s)$ for each case so that it has an integrating pole at the origin if possible.

$$(a) \quad T(s) = \frac{K_c(s+40)}{s^2 + 50s + 625} \quad ; \quad H(s) = K_h$$

$$(b) \quad T(s) = \frac{K_c}{s(s+30)} \quad ; \quad H(s) = 5s$$

$$(c) \quad T(s) = \frac{K_c(s+20)}{(s+0.1)(s^2+130s+150^2)} \quad ; \quad H(s) = \frac{s}{s+20}$$

11-7. Use the direct root locus method to set K in Prob. 9-10 so that the closed-loop poles have a damping ratio of 0.5. What is the system velocity constant?

11-8. Use the direct root locus method to determine K in Prob. 9-11 for a dominant pole damping ratio of 0.6.

11-9. Draw the direct root locus for a lag compensated modification of the system of Prob. 9-11. The system K_v is to be increased 5 times with the damping ratio kept at 0.707 for the complex pair of closed-loop poles.

11-10. Repeat Prob. 9-14 except design so that the dominant closed-loop poles have a damping ratio of 0.6 if possible. Use the second-order system curves of Chapter 6 or Chapter 13 to obtain an approximate overshoot figure. Determine the final system bandwidth by using a Bode plot of the closed-loop transfer function.

11-11. Using the direct root locus, adjust the gain of the system of Prob. 9-17 before compensation for $\zeta = 0.5$. Redraw the direct root locus including the lead compensation pole and zero given in Prob. 9-17. How is the locus modified?

11-12. A unity feedback system with $KG(s) = K/(s)^2(s + 20)$ is to be lead-compensated to make it stable so as to have an overshoot of 25 percent or less to a unit-step input. Construct the original and the compensated root locus. Is the final system desirable if the gain changes?

11-13. Determine the maximum overshoot and the time at which it occurs for

$$c(t) = 1 + \frac{p_1 q_z}{z_1 \rho_p \sin \theta} e^{-at} \sin(\beta t - \pi + \theta - \varphi_p + \varphi_z)$$

This $c(t)$ is an approximate version of the actual $c(t)$ resulting from a pole-zero pattern of two dominant complex poles ($-a \pm j\beta = \rho e^{\pm j(\alpha - \theta)}$), one real pole ($-p_1$), and one real zero ($-z_1$). The exponential term $A_1 e^{-p_1 t}$ has been dropped. The terms q_z and ρ_p are phasor distances from the zero at $-z_1$ and the pole at $-p_1$ to the pole at $-a + j\beta$ while the angles φ_z and φ_p are the angles of these phasors. This scheme gives a reasonably accurate but simple method of finding overshoot for higher-order systems. Additional poles and zeros can easily be added. What effect do added poles have on the overshoot? What effect do added zeros have on the overshoot?

11-14. The cyclical process of the direct and indirect root loci is to be used to investigate the possibility of compensating a system with $KG_2(s) = K_2/(s)(s + 5)$ so as to meet the following specifications:

$$\begin{aligned} K_v &\geq 50 \\ \text{Overshoot} &\leq 10\% \\ \text{Rise time} &\leq 0.2 \text{ sec} \\ BW &\leq 20 \text{ radian/sec} \end{aligned}$$

Consider the uncompensated system first and set K_2 for 10 per cent overshoot (use the second-order system curves of Fig. 13-2-1.) Determine what quantities must be changed. Choose a new set of closed-loop poles and zeros that give the right K_v , approximate BW (use a Bode plot if necessary), and overshoot (use the approximation of Prob. 11-13 if valid). Can the rise time be met now? Construct the inverse root locus and determine the open-loop transfer function such that the pole at the origin in the original $KG_2(s)$ is included. What compensation is required?

11-15. Repeat Prob. 11-14 except use

$$KG_2(s) = \frac{K_2}{s(20s + 1)(100s + 1)}$$

$$\begin{aligned} K_v &\geq 0.5 \\ \text{Overshoot} &\leq 20\% \\ BW &\leq 1.0 \text{ radians/sec} \end{aligned}$$

11-16. Repeat Prob. 11-14 using the same system and specifications except use one more cycle of the direct-indirect root locus process and try to simplify the compensation.

11-17. In Ex. 7-7-3, the Nyquist diagram is used on

$$KG(s) = \frac{1}{s(s + 1)} \quad ; \quad H(s) = 1$$

to reason that the poles of the closed-loop system are located on a line $s = -0.5 + j\omega$. Use the direct root locus to verify this result.

11-18. In Ex. 7-7-4, the Nyquist diagram is used on

$$KG(s) = \frac{K(s+4)}{s(s-1)} \quad ; \quad H(s) = 1$$

and the conclusions are reached that if K is above unity the closed-loop system is stable. Use the direct root locus to verify this result.

11-19. Chapter 7 contains many problems that are to be solved by using Nyquist diagrams. These same problems are to be reworked by using root locus methods as assigned in the following parts:

- | | | |
|----------------|----------------|----------------|
| (a) Prob. 7-12 | (f) Prob. 7-19 | (k) Prob. 7-24 |
| (b) Prob. 7-13 | (g) Prob. 7-20 | (l) Prob. 7-25 |
| (c) Prob. 7-14 | (h) Prob. 7-21 | (m) Prob. 7-26 |
| (d) Prob. 7-15 | (i) Prob. 7-22 | (n) Prob. 7-28 |
| (e) Prob. 7-17 | (j) Prob. 7-23 | (o) Prob. 7-29 |

12

NETWORK SYNTHESIS

12-1. Introduction. Network theory can be divided into two categories. The more familiar category is analysis, in which the network is known and the problem is to find the response of the network in a certain situation. The second category is synthesis, in which certain properties of the network are specified and the problem is to determine the network.

Synthesis, in turn, is composed of two divisions. The first of these is referred to as the approximation problem. The network to be synthesized may be specified indirectly. For example, the input and the output functions may be given graphically. From these graphical data, a transfer function must be determined such that the data are approximately met but at the same time the transfer function must be such that it is physically realizable.

The second division is the realizability procedures. After a transfer function is determined, it is necessary to find a network that yields such a function. In general, not only one network can be found but an infinite number of networks are possible.

In control systems, the network to be synthesized is for use at low frequencies. In most situations it is impracticable to use inductors because of the weight, size, and cost limitations. Therefore control system theory is primarily concerned with *RC* networks. However, certain *RC* synthesis procedures are closely related to *LC* procedures; therefore, both types of networks are considered in parts of this discussion.

12-2. Poles and Zeros of *LC* Driving-Point Impedance. An *LC* network has the operational impedance

$$Z(s) = \frac{E(s)}{I(s)} = \frac{N(s)}{D(s)} \quad (12-2-1)$$

The network can be excited by an impulse of current and the transformed response is

$$E(s) = Z(s) = \frac{N(s)}{D(s)} \quad (12-2-2)$$

The network can be excited by an impulse of voltage and the transformed response is

$$I(s) = \frac{1}{Z(s)} = \frac{D(s)}{N(s)} \quad (12-2-3)$$

When the network is driven by a current source, the terminals of the network are open-circuited since the internal impedance of a current source is infinite. Therefore, the poles of $Z(s)$ can be referred to as the natural frequencies of the network with its terminals open-circuited. When the network is driven by a voltage source, the terminals of the network are short-circuited since the internal impedance of a voltage source is zero. Therefore the zeros of $Z(s)$ can be termed the natural frequencies of the network with its terminals short-circuited.

Since the network contains no resistors, it will oscillate indefinitely after it is excited by either a current source or a voltage source. Therefore all the poles and zeros of $Z(s)$ must be on the $j\omega$ -axis. These oscillations do not build up; therefore, it can be reasoned that the poles and zeros are simple.

Since the network is composed of only reactances, the phase angle of the impedance in the a-c circuit sense is either 90° or -90° . This phase angle is the angle that results from $Z(j\omega)$ as ω moves up the $j\omega$ -axis. It can be reasoned that the $Z(j\omega)$ phase angle can remain either 90° or -90° only if the poles and zeros are interlaced.¹

The physical realizability conditions for an LC driving-point impedance (reactance function) are established. These conditions are that the poles and zeros must be simple and be on the $j\omega$ -axis with the poles and zeros interlaced.

12-3. Poles and Zeros of RC Driving-Point Impedance. Eqs. 12-2-1, 12-2-2, and 12-2-3 apply to RC as well as to LC networks. Since the network does not contain any inductors, there can be no exchange of energy between inductors and capacitors and hence the response is non-oscillatory when the network is excited by either a current or voltage source. Therefore all the poles and zeros of $Z(s)$ must be on the negative real axis.

The relative locations among the poles and zeros can be established by going to sinusoidal steady-state conditions and reasoning that the angle on the phasor impedance must lie between 0° and -90° . If a number of poles are grouped together on the negative real axis, and a number of zeros are grouped together farther out on the axis, the net phase on $Z(j\omega)$ can certainly exceed -90° . It can intuitively be seen

¹ This result can best be visualized by assuming a small amount of dissipation so that the poles and zeros are slightly to the left of the $j\omega$ -axis. If two zeros (or two poles) are adjacent, the total phase angle falls outside the range of $\pm 90^\circ$ as ω moves up the $j\omega$ -axis in the vicinity of the two zeros (or poles).

that for the phase angle on $Z(j\omega)$ to have a value in the range of 0° to -90° for all frequencies, the poles and zeros must be simple and be interlaced. In addition, since the total phase angle is negative, the critical frequency (a term used to mean either a pole or a zero) nearest the origin must be a pole and the highest critical frequency must be a zero.

The physical realizability conditions for an RC driving-point impedance are established. These conditions are that the poles and zeros be simple and lie on the negative real axis with the poles and zeros interlaced. In addition to this, the lowest critical frequency is a pole and the highest critical frequency is a zero.

12-4. Poles and Zeros of RL Driving-Point Impedance. RL networks are little used in control systems; however, their physical realizability conditions can be determined in a manner similar to that used for RC networks. These conditions are that the poles and zeros be simple and on the negative real axis with the poles and zeros interlaced. In addition to this, the lowest critical frequency is a zero and the highest critical frequency is a pole.

It can be seen that a function that can be realized as an RC driving-point impedance can also be realized as an RL driving-point admittance, and vice versa.

12-5. L - C Realization Methods. The basic philosophy of any synthesis method is removal of a physically realizable element from a network in such a manner that the remaining network is also physically realizable and in general is simpler.

LC networks can be synthesized in any one of four canonic forms. A canonic form is a network that realizes an impedance function with a minimum number of elements. As is discussed later, these canonic forms can be mixed in any one synthesis problem. The four canonic forms are Foster's first and second forms and Cauer's first and second forms.

An example is used to illustrate all four forms. As discussed in Chapter 7, if a polynomial is Hurwitz, the ratio of the even part to the odd part or the odd part to the even part is a reactance function. The polynomial of Ex. 7-3-2 is used. The polynomial is repeated as

$$s^5 + 42s^4 + 601s^3 + 3552s^2 + 7220s + 11600 \quad (12-5-1)$$

An LC driving point impedance is formed as

$$Z(s) = \frac{s^5 + 601s^3 + 7220s}{42s^4 + 3552s^2 + 11600} \quad (12-5-2)$$

FOSTER'S FIRST FORM. The first Foster form expands $Z(s)$ into partial fractions as

$$\begin{aligned}
 Z(s) &= \frac{s^3 + 601s^2 + 7220s}{42(s^2 + 3.4)(s^2 + 81.1)} \\
 &= \frac{s^3 + 601s^2 + 7220s}{42(s - j1.845)(s + j1.845)(s - j9.0)(s + j9.0)} \\
 &= \frac{s}{42} + \frac{0.795}{s - j1.845} + \frac{0.795}{s + j1.845} + \frac{5.35}{s - j9.0} + \frac{5.35}{s + j9.0} \\
 &= \frac{s}{42} + \frac{1.59s}{s^2 + 3.4} + \frac{10.7s}{s^2 + 81.1} \quad (12-5-3)
 \end{aligned}$$

Since these three terms represent an impedance, the corresponding network must consist of three groups of elements connected in series. In order to recognize the second and third terms of the expansion, the operational impedance of the parallel LC circuit of Fig. 12-5-1 is determined:

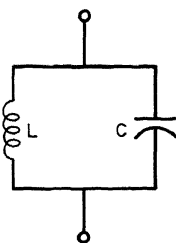


Fig. 12-5-1.

$$Z(s) = \frac{sL \times (1/sC)}{sL + 1/sC} = \frac{(1/C)s}{s^2 + 1/LC} \quad (12-5-4)$$

The second term is realized as an LC parallel combination to demonstrate the procedure. This is done by equating

$$\frac{(1/C)s}{s^2 + 1/LC} = \frac{1.59s}{s^2 + 3.4} \quad (12-5-5)$$

from which
$$\frac{1}{C} = 1.59$$

and
$$\frac{1}{LC} = 3.4$$

From these equations, the values of L and C are found as

$$L = 0.468 \quad C = 0.628 \quad (12-5-6)$$

An alternate method of finding L and C is to note that at low frequencies an LC parallel combination acts as if it were an inductor, and for small values of s the second term of the expansion in Eq. 12-5-3 becomes

$$Ls = \frac{1.59}{3.4} s$$

from which

$$L = 0.468$$

At high frequencies an LC parallel combination acts as if it were a capacitor, and for large values of s the second term becomes

$$\frac{1}{Cs} = \frac{1.59}{s}$$

from which

$$C = 0.628$$

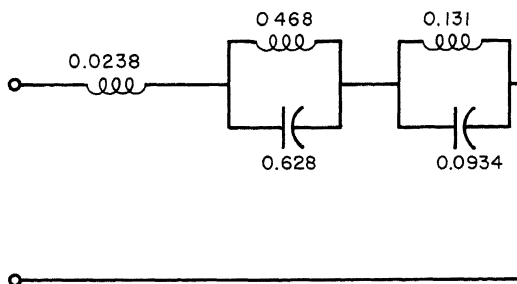


Fig. 12-5-2.

The third term can be evaluated in the same manner and the network that realizes the LC function is shown in Fig. 12-5-2.

FOSTER'S SECOND FORM. Foster's second form expands $Y(s)$ in partial fractions as

$$\begin{aligned} Y(s) &= \frac{42s^4 + 3552s^2 + 11600}{s(s^2 + 12.2)(s^2 + 589)} \\ &= \frac{42s^4 + 3552s^2 + 11600}{s(s - j3.94)(s + j3.94)(s - j24.3)(s + j24.3)} \\ &= \frac{1.607}{s} + \frac{1.82}{s - j3.94} + \frac{1.82}{s + j3.94} + \frac{18.38}{s - j24.3} + \frac{18.38}{s + j24.3} \\ &= \frac{1.607}{s} + \frac{3.64s}{s^2 + 12.2} + \frac{36.76s}{s^2 + 589} \end{aligned} \quad (12-5-7)$$

Since these three terms represent an admittance, the corresponding network must consist of three groups of elements connected in parallel. To recognize the second and third terms, the operational admittance for the series LC current of Fig. 12-5-3 is determined:

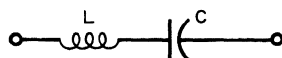


Fig. 12-5-3.

$$Y(s) = \frac{1}{Ls + 1/Cs} = \frac{(1/L)s}{s^2 + 1/LC} \quad (12-5-8)$$

The second term is realized as an LC series combination to demonstrate the procedure. The second term is

$$\frac{3.64s}{s^2 + 12.2} \quad (12-5-9)$$

At low frequencies a series LC combination acts as if it were a capacitor. For small values of s , this second term becomes

$$Cs = \frac{3.64}{12.2} s \quad (12-5-10)$$

from which

$$C = 0.298 \quad (12-5-11)$$

At high frequencies a series LC combination acts as if it were inductor. For large values of s , this second term becomes

$$\frac{1}{Ls} = \frac{3.64}{s} \quad (12-5-12)$$

from which

$$L = 0.275 \quad (12-5-13)$$

The third term in the expansion can be evaluated in the same manner and the network that realizes the LC function is shown in Fig. 12-5-4.

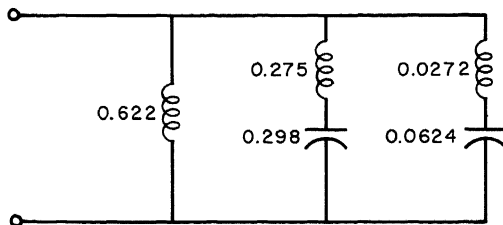


Fig. 12-5-4.

CAUER'S FIRST FORM. The first Cauer form removes the pole at infinity. If $Z(s)$ has a pole at infinity, the pole is removed. However, if $Z(s)$ has a zero at infinity the function is "flipped over" and the pole is removed from the admittance function. In the example,

$$Z(s) = \frac{s^5 + 601s^3 + 7220s}{42s^4 + 3552s^2 + 11,600} \quad (12-5-14)$$

The impedance function has the pole at infinity. Therefore the procedure begins by removing this pole.

For large values of s , $Z(s)$ behaves as

$$Z(s) \Big|_{s \rightarrow \infty} = \frac{s^5}{42s^4} = \frac{s}{42} \quad (12-5-15)$$

To remove the pole at infinity, $s/42$ is subtracted from $Z(s)$ and the remainder function is called $Z_1(s)$.

$$Z_1(s) = \frac{s^5 + 601s^3 + 7220s}{42s^4 + 3552s^2 + 11,600} - \frac{s}{42} = \frac{516s^3 + 6940s}{42s^4 + 3552s^2 + 11,600} \quad (12-5-16)$$

$Z_1(s)$ has a zero at infinity, therefore $Z_1(s)$ is "flipped over" and the pole at infinity in $Y_1(s)$ is removed leaving $Y_2(s)$:

$$Y_2(s) = \frac{42s^4 + 3552s^2 + 11,600}{516s^3 + 6944s} - \frac{42s}{516} = \frac{2980s^2 + 11,600}{516s^3 + 6940s} \quad (12-5-17)$$

$Y_2(s)$ has a zero at infinity, therefore $Y_2(s)$ is "flipped over" and the pole at infinity in $Z_2(s)$ is removed leaving $Z_3(s)$:

$$Z_3(s) = \frac{516s^3 + 6940s}{2980s^2 + 11,600} - \frac{516s}{2980} = \frac{4900s}{2980s^2 + 11,600} \quad (12-5-18)$$

$Z_3(s)$ is of such a form that it can be recognized as a parallel LC combination. However, the method of removing the pole at infinity is continued:

$$Y_4(s) = \frac{2980s^2 + 11,600}{4900s} - \frac{2980s}{4900} = \frac{2.37}{s} \quad (12-5-19)$$

When all of these terms are collected, the following expansion has been made on $Z(s)$:

$$Z(s) = 0.0238s + \frac{1}{0.0814s + \frac{1}{0.173s + \frac{1}{0.608s + \frac{1}{0.422s}}}} \quad (12-5-20)$$

This type of expansion is known as the continued-fraction expansion. Since it represents an impedance function, the $0.0238s$ term is an inductor $L = 0.0238$ in series with the remainder of the network. The impedance term on the right is equal to 1 over an admittance function. The admittance is made up of a capacitor $C = 0.0814$ in parallel with the remainder of the network. The admittance is 1 over an impedance function which is made up of an inductor $L = 0.173$ in series with the rest of the network. This process is continued until the network of Fig. 12-5-5 is obtained.

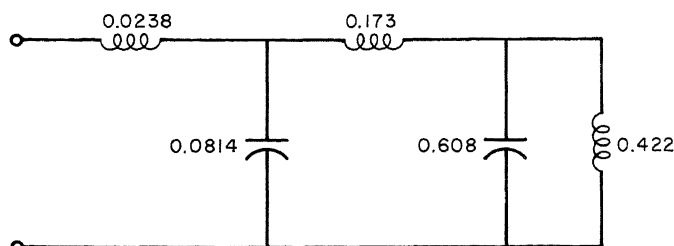


Fig. 12-5-5.

The polynomial of Eq. 12-5-1 is used to obtain the reactance function for these examples. It is of interest to develop the Routh table for this polynomial. For convenience, the polynomial is repeated as

$$s^5 + 42s^4 + 601s^3 + 3552s^2 + 7220s + 11,600 \quad (12-5-21)$$

The Routh table is given as

s^5	1	601	7,220
s^4	42	3,552	11,600
s^3	516	6,940	
s^2	2,980	11,600	
s^1	4,900		
s^0	11,600		

(12-5-22)

The ratio of the first row to the second row is the reactance function synthesized in these examples. The ratio of the second row to the third row is $Y_1(s)$, as can be seen by comparing the coefficients in the table with the reciprocal of Eq. 12-5-16. The ratio of the third row to the fourth row is $Z_2(s)$, as can be seen from the reciprocal of Eq. 12-5-17. The ratio of the fourth row to the fifth row is $Y_3(s)$, and finally the ratio of the fifth row to the sixth row is $Z_4(s)$. Therefore it can be seen that the Routh table is another method of obtaining a network synthesized by Cauer's first form.

The coefficients in the continued fraction expansion of Eq. 12-5-20 can be obtained from the ratio of successive numbers in the first column of the Routh table. These ratios are

$$1/42 = 0.0238, \quad 42/516 = 0.0814, \quad 516/2980 = 0.173, \text{ etc.}$$

CAUER'S SECOND FORM. The second Cauer form removes the pole at the origin. If $Z(s)$ has a pole at the origin, the pole is removed. However, if $Z(s)$ has a zero at the origin, the function is "flipped over" and the pole is removed from the admittance function.

The substitution

$$s = \frac{1}{z} \quad (12-5-23)$$

exchanges the roles of the origin and infinity. With this substitution, the polynomial

$$s^5 + 42s^4 + 601s^3 + 3552s^2 + 7220s + 11,600 \quad (12-5-24)$$

becomes

$$11,600z^5 + 7220z^4 + 3552z^3 + 601z^2 + 42z + 1 \quad (12-5-25)$$

The Routh table is formed as

z^5	11,600	3,552	42
z^4	7,220	601	1
z^3	2,585	40.39	
z^2	488.6	1	
z^1	35.1		
z^0	1		

(12-5-26)

The continued-fraction expansion is obtained from the ratio of successive numbers in the first column and the substitution $z = 1/s$ is again used. An examination of the method in which this expansion is obtained indicates that it represents an admittance function:

$$Y(s) = \frac{1.61}{s} + \frac{1}{\frac{2.79}{s} + \frac{1}{\frac{5.3}{s} + \frac{1}{\frac{13.9}{s} + \frac{1}{\frac{35.1}{s}}}}} \quad (12-5-27)$$

The network of Fig. 12-5-6 is obtained.

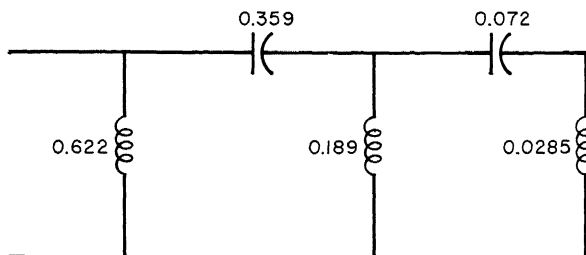


Fig. 12-5-6.

DISCUSSION. Each of the networks obtained by the four canonic forms contains five elements. However, there are more than four networks that can synthesize this function with five elements, because the canonic forms can be mixed. For example the series inductor can be removed by Foster's first form, and the synthesis procedure then switched to Foster's second form to remove a shunt LC series combination, etc.

It is also possible to synthesize this same impedance function with an infinite number of networks each containing an arbitrary number of elements. For example, each inductor and capacitor can be divided into

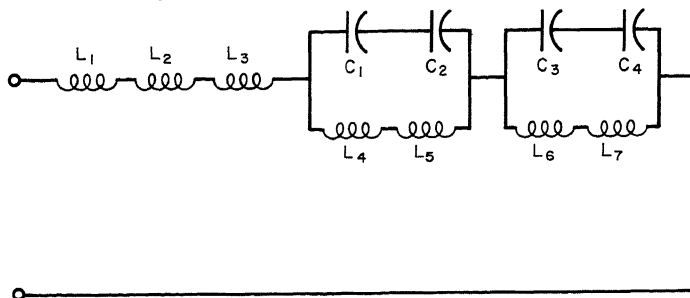


Fig. 12-5-7.

a group of series elements, as indicated in Fig. 12-5-7. This example may seem as if it is a trivial case; however, this same type of division of elements can be disguised by removing L_1 in Fig. 12-5-7, and then expanding the remainder function in another form such as Foster's second form. The residue at one of the finite poles can be calculated but instead of removing all the pole, only a part of the pole is removed. After this partial pole removal, the network can be "flipped over" and a part of a finite pole in Foster's first form can be removed. The network obtained

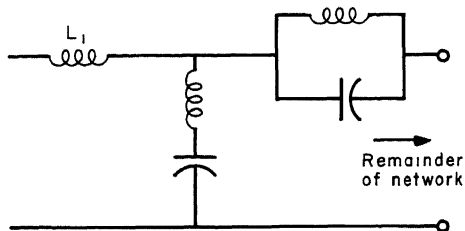


Fig. 12-5-8.

by these steps is indicated in Fig. 12-5-8. The function describing the remainder of the network is as complicated as the original function because none of the poles has been removed. This type of expansion can be continued indefinitely.

12-6. RC Realization Methods. RC networks can be synthesized in any one of four canonic forms, which are analogous to the LC forms and for convenience are referred to by the same names.

FOSTER'S FIRST FORM. Foster's first form expands the impedance function $Z(s)$ into partial fractions. A $Z(s)$ and its expansion are given as

$$\begin{aligned} Z(s) &= \frac{(s+1)(s+3)}{s(s+2)(s+4)} \\ &= \frac{0.375}{s} + \frac{0.250}{s+2} + \frac{0.375}{s+4} \end{aligned} \quad (12-6-1)$$

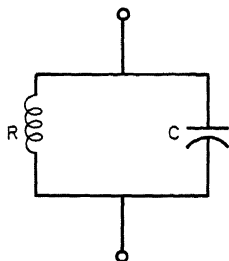


Fig. 12-6-1.

In order to recognize the second and third terms of this expansion, the operational impedance of the parallel RC circuit of Fig. 12-6-1 is determined.

$$Z(s) = \frac{R \times \frac{1}{Cs}}{R + \frac{1}{Cs}} = \frac{\frac{1}{C}}{s + \frac{1}{RC}} \quad (12-6-2)$$

The second term of Eq. 12-6-1 is used to demonstrate the manner of determining the R 's and C 's. At low frequencies this RC combination acts as a resistor, and for small values of s the second term becomes

$$R = \frac{0.250}{2} = 0.125 \quad (12-6-3)$$

At high frequencies, this RC combination acts as a capacitor, and for large values of s the second term becomes

$$\frac{1}{Cs} = \frac{0.250}{s} \quad (12-6-4)$$

from which C is found as

$$C = 4 \quad (12-6-5)$$

The third term can be evaluated in a similar manner and the network that realizes the RC function is shown in Fig. 12-6-2.

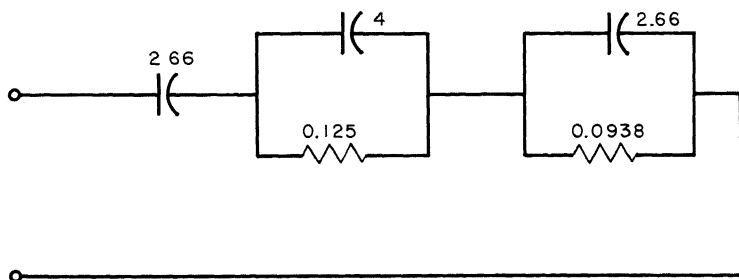


Fig. 12-6-2.

This discussion can be generalized to demonstrate several properties of RC networks. Suppose that a $Z(s)$ and its expansion are

$$Z(s) = \frac{(s + z_1)(s + z_2)(s + z_3)}{(s + p_1)(s + p_2)(s + p_3)}$$

$$= 1 + \frac{K_{-p_1}}{s + p_1} + \frac{K_{-p_2}}{s + p_2} + \frac{K_{-p_3}}{s + p_3} \quad (12-6-6)$$

Where $p_1 < z_1 < p_2 < z_2 < p_3 < z_3$. Each of the K 's is positive, as can be seen from the following discussion. The residue K_{-p_1} is obtained by "covering up" the $s + p_1$ factor and substituting $s = -p_1$ in the remainder of the function. Since p_1 is smaller than any of the other p 's and z 's, all the factors that make up K_{-p_1} are positive, and therefore K_{-p_1} is positive. K_{-p_2} is obtained by "covering up" the $s + p_2$ factor and substituting $s = -p_2$ in the remainder of the function. Since p_2 is larger than p_1 and z_1 , two of the factors that make up K_{-p_2} are negative, with the result that K_{-p_2} is positive. By a similar line of reasoning, it can be seen that K_{-p_3} is also positive.

If s in Eq. 12-6-6 is restricted to its real values $s = \sigma$, the equation becomes

$$Z(\sigma) = 1 + \frac{K_{-p_1}}{\sigma + p_1} + \frac{K_{-p_2}}{\sigma + p_2} + \frac{K_{-p_3}}{\sigma + p_3} \quad (12-6-7)$$

When $Z(\sigma)$ is differentiated with respect to σ , the result is

$$\frac{d[Z(\sigma)]}{d\sigma} = -\frac{K_{-p_1}}{(\sigma + p_1)^2} - \frac{K_{-p_2}}{(\sigma + p_2)^2} - \frac{K_{-p_3}}{(\sigma + p_3)^2} \quad (12-6-8)$$

From this it can be seen that the derivative of $Z(\sigma)$ with respect to σ is negative for all values of σ . A plot of the $Z(\sigma)$ of Eq. 12-6-6 is shown in Fig. 12-6-3.

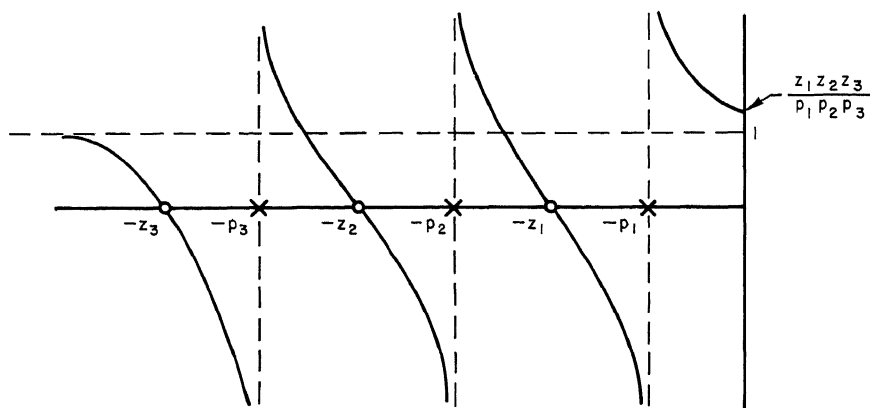


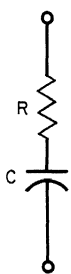
Fig. 12-6-3.

Another property of RC networks is that the magnitude of the impedance at $s = 0$ is always greater than the magnitude at $s = \infty$. This statement is true because all the capacitors become short circuits at $s = \infty$.

FOSTER'S SECOND FORM. Foster's second form expands the admittance functions in partial fractions. If this is done on $Y(s)$ without any preliminary step, all the residue terms are negative and it is impossible to recognize a network. In order to explore the type of expansion that is desired, the operational admittance of the RC series circuit of Fig. 12-6-4 is found as

$$Y(s) = \frac{1}{R + 1/sC} = \frac{(1/R)s}{s + 1/RC} \quad (12-6-9)$$

Therefore $Y(s)$ must be expanded so that the s factor appears in the numerator of each term. To do this, $Y(s)/s$ is expanded and the result multiplied by s to yield $Y(s)$. This is done for the example of Eq. 12-6-1:



$$\frac{Y(s)}{s} = \frac{(s+2)(s+4)}{(s+1)(s+3)} = 1 + \frac{1.5}{s+1} + \frac{0.5}{s+3} \quad (12-6-10)$$

from which $Y(s)$ is found to be

$$Y(s) = s + \frac{1.5s}{s+1} + \frac{0.5s}{s+3} \quad (12-6-11)$$

Fig. 12-6-4.

The second term of Eq. 12-6-11 is used to demonstrate the method of determining the R 's and C 's. At low frequencies the series RC circuit acts as a capacitor, and for small values of s the second term becomes

$$Cs = 1.5s$$

from which

$$C = 1.5 \quad (12-6-12)$$

At high frequencies this RC combination acts as a resistor, and for large values of s the second term becomes

$$\frac{1}{R} = 1.5$$

from which

$$R = 0.667 \quad (12-6-13)$$

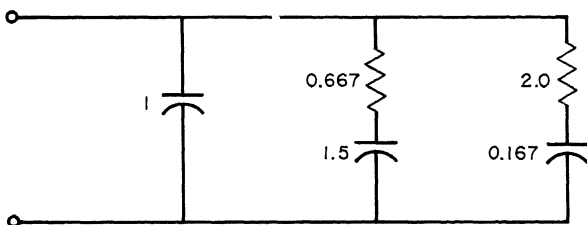


Fig. 12-6-5.

The third term can be evaluated in a similar manner and the network that realizes the RC function is shown in Fig. 12-6-5.

CAUER'S FIRST FORM. Cauer's first form removes the pole at infinity. An RC impedance can never have a pole at infinity but it may have a zero at infinity. If it does, the function is "flipped over" and the pole at infinity in the admittance is removed. If the impedance does not have a zero at infinity, a preliminary step is required to create this zero.

The previous example is used again:

$$Z(s) = \frac{(s+1)(s+3)}{s(s+2)(s+4)} = \frac{s^2 + 4s + 3}{s^3 + 6s^2 + 8s} \quad (12-6-14)$$

$Z(s)$ has a zero at infinity; therefore, it is "flipped over" and the pole at infinity in $Y(s)$ is removed leaving $Y_1(s)$.

$$Y_1(s) = \frac{s^3 + 6s^2 + 8s}{s^2 + 4s + 3} - s = \frac{2s^2 + 5s}{s^2 + 4s + 3} \quad (12-6-15)$$

To continue the process, $Y_1(s)$ is "flipped over" and the constant value at infinity is removed to create a zero at infinity in $Z_2(s)$:

$$Z_2(s) = \frac{s^2 + 4s + 3}{2s^2 + 5s} - 0.5 = \frac{1.5s + 3}{2s^2 + 5s} \quad (12-6-16)$$

$Z_2(s)$ is "flipped over" and the pole at infinity is removed leaving $Y_3(s)$:

$$Y_3(s) = \frac{2s^2 + 5s}{1.5s + 3} - 1.33s = \frac{s}{1.5s + 3} \quad (12-6-17)$$

Although $Y_3(s)$ can be recognized as a series RC combination, the procedure is continued by "flipped over" $Y_3(s)$ and removing the constant value at infinity and leaving $Z_4(s)$

$$Z_4(s) = \frac{1.5s + 3}{s} - 1.5 = \frac{3}{s} \quad (12-6-18)$$

The above work can be condensed into a division process:

$$s^2 + 4s + 3 \overline{) \begin{array}{l} s^3 + 6s^2 + 8s \\ s^3 + 4s^2 + 3s \\ \hline 2s^2 + 5s \end{array}} \quad (12-6-19)$$

$$\begin{array}{r} \begin{array}{l} s^2 + 4s + 3 \\ \hline s^2 + 2.5s \\ \hline 1.5s + 3 \end{array} \quad \begin{array}{l} (1.33s) \\ \hline 2s^2 + 5s \\ \hline 2s^2 + 4s \\ \hline s \end{array} \quad \begin{array}{l} (1.5) \\ \hline 1.5s + 3 \\ \hline 1.5s \\ \hline 3 \end{array} \quad \begin{array}{l} (s/3) \\ \hline s \\ \hline 0 \end{array} \end{array}$$

The result can be put into the continued-fraction form as

$$Y(s) = s + \frac{1}{0.5 + \frac{1}{1.33s + \frac{1}{1.5 + \frac{1}{\frac{s}{3}}}}} \quad (12-6-20)$$

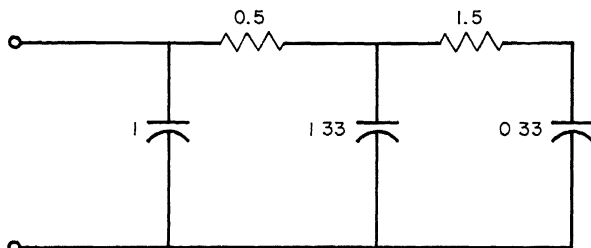


Fig. 12-6-6.

The resulting network is shown in Fig. 12-6-6.

An examination of the division in Eq. 12-6-19 shows that the first division removes the pole at infinity in $Y(s)$ and the remainder divided by $s^2 + 4s + 3$ is $Y_1(s)$. $Y_1(s)$ is "flipped over" by the manner in which the next division is carried out and the constant term at infinity is removed. The remainder divided by $2s^2 + 5s$ is $Z_2(s)$. $Z_2(s)$ is "flipped over" by the manner in which the next division is carried out, etc.

CAUER'S SECOND FORM. Cauer's second form removes the pole at zero. An RC impedance may or may not have a pole at zero. If it has a such pole, this pole is removed. If it does not, then $Z(s)$ is flipped over," and $Y_1(s)$ is created with a zero at $s = 0$ by removing the constant value in $Y(s)$ at zero. This procedure is repeated until the network is synthesized.

The previous example is used again:

$$Z(s) = \frac{s^2 + 4s + 3}{s^3 + 6s^2 + 8s} \quad (12-6-21)$$

Since $Z(s)$ has a pole at zero, no preliminary step is needed before this pole can be removed. The entire procedure is condensed into a division process similar to that in Eq. 12-6-19:

$$\begin{array}{r} (0.375/s) \\ 8s + 6s^2 + s^3 \overline{) 3 + 4s + s^2} \\ \underline{3 + 2.25s + 0.375s^2} \quad (4.57) \\ 1.75s + 0.625s^2 \overline{) 8s + 6s^2 + s^3} \\ \underline{8s + 2.86s^2} \quad (0.557/s) \\ 3.14s^2 + s^3 \overline{) 1.75s + 0.625s^2} \\ \underline{1.75s + 0.557s^2} \quad (46) \\ 0.068s^2 \overline{) 3.14s^2 + s^3} \\ \underline{3.14s^2} \quad (0.068/s) \\ s^3 \overline{) 0.068s^2} \\ \underline{0.068s^2} \\ 0 \end{array} \quad (12-6-22)$$

The result can be put into the continued-fraction form as

$$Z(s) = \frac{0.375}{s} + \frac{1}{4.57 + \frac{1}{\frac{0.557}{s} + \frac{1}{46 + \frac{1}{\frac{0.068}{s}}}}} \quad (12-6-23)$$

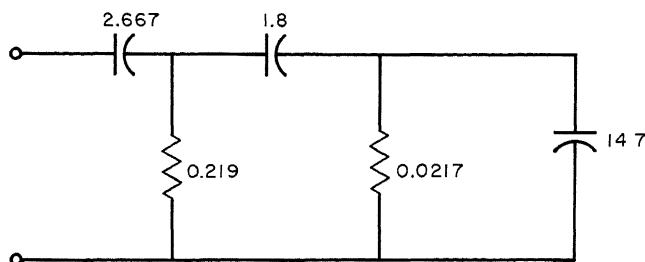


Fig. 12-6-7.

The resulting network is shown in Fig. 12-6-7.

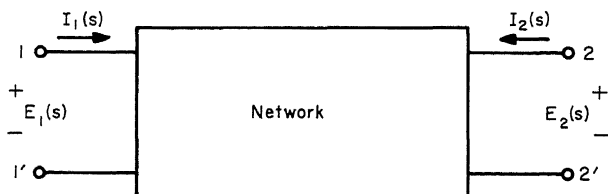


Fig. 12-7-1.

12-7. Two-Terminal-Pair Networks. A two-terminal-pair network is shown in Fig. 12-7-1. Of the several sets of parameters available to describe this network, two sets are featured for the present purposes.

The first set of parameters is defined by the equations

$$E_1 = z_{11}I_1 + z_{12}I_2 \quad (12-7-1)$$

$$E_2 = z_{21}I_1 + z_{22}I_2 \quad (12-7-2)$$

Each of these voltages, currents, and z 's is a function of s , but the s is dropped for convenience in writing.

If I_2 in Eq. 12-7-1 is zero, the resulting equation can be solved for z_{11} as

$$z_{11} = \frac{E_1}{I_1} \quad (12-7-3)$$

Therefore z_{11} is the driving-point impedance of the 1-1' terminals, with the 2-2' terminals left open.

In a similar manner, it can be seen that z_{22} is the driving-point impedance of the 2-2' terminals with the 1-1' terminals left open.

If I_1 in Eq. 12-7-1 is zero, the resulting equation can be solved for z_{12}

$$z_{12} = \frac{E_1}{I_2} \quad (12-7-4)$$

and in a similar manner, z_{21} can be found to be

$$z_{21} = \frac{E_2}{I_1} \quad (12-7-5)$$

As a result of the reciprocity theorem,

$$z_{12} = z_{21} \quad (12-7-6)$$

The parameter z_{12} is a transfer impedance as the current is supplied at one set of terminals and the resulting voltage is found at the other set.

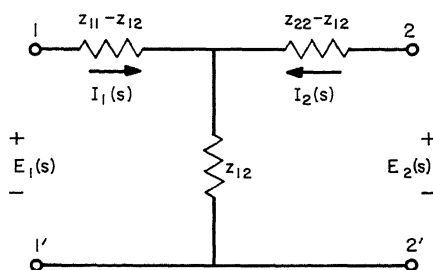


Fig. 12-7-2.

The network of Fig. 12-7-1 can be represented by an equivalent T network as shown in Fig. 12-7-2. The driving-point impedance at the 1-1' terminals with the 2-2' terminals open is z_{11} and that at the 2-2' terminals with the 1-1' terminals open is z_{22} . Also, the network can be driven by a current I_2 and the resulting voltage is

$$E_1 = I_2 z_{12} \quad (12-7-7)$$

or by a current I_1 and the resulting voltage is

$$E_2 = I_1 z_{12} \quad (12-7-8)$$

Therefore, this network is the equivalent circuit.

The second set of parameters is defined by the equations

$$I_1 = y_{11}E_1 + y_{12}E_2 \quad (12-7-9)$$

$$I_2 = y_{21}E_1 + y_{22}E_2 \quad (12-7-10)$$

If E_2 in Eq. 12-7-9 is zero, the resulting equations can be solved for y_{11}

$$y_{11} = \frac{I_1}{E_1} \quad (12-7-11)$$

Therefore y_{11} is the driving-point admittance of the 1-1' terminals with the 2-2' terminals shorted.

In a similar manner, it can be seen that y_{22} is the driving-point admittance of the 2-2' terminals with the 1-1' terminals shorted.

If E_1 in Eq. 12-7-9 is zero, then the resulting equation can be solved for y_{12} as

$$y_{12} = \frac{I_1}{E_2} \quad (12-7-12)$$

and in a similar manner y_{21} can be found as

$$y_{21} = \frac{I_2}{E_1} \quad (12-7-13)$$

As a result of the reciprocity theorem,

$$y_{12} = y_{21} \quad (12-7-14)$$

The parameter y_{12} is a transfer admittance because the voltage is supplied at one set of terminals and the resulting current is found at the other set.

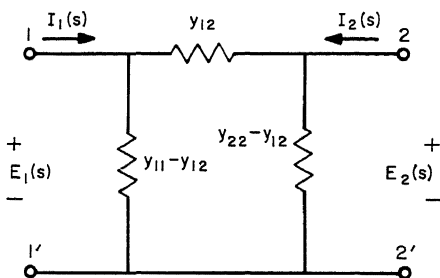


Fig. 12-7-3.

The network of Fig. 12-7-1 can be represented by an equivalent π network as shown in Fig. 12-7-3. The fact that this is the equivalent circuit can be reasoned in a manner similar to that for the equivalent T.

The z 's and the y 's are not reciprocals because the z 's are obtained with the other pair of terminals opened and the y 's are obtained with the other pair of terminals shorted.

To find the relationship between the y 's and z 's, the 2-2' terminals of Fig. 12-7-2 can be shorted and the admittance at the 1-1' terminals becomes y_{11} . An equation can be written for this circuit, and y_{11} can be found to be

$$y_{11} = \frac{z_{22}}{z_{11}z_{22} - z_{12}^2} \quad (12-7-15)$$

In a similar manner, y_{22} and y_{12} can be found to be

$$y_{22} = \frac{z_{11}}{z_{11}z_{22} - z_{12}^2} \quad (12-7-16)$$

and

$$y_{12} = -\frac{z_{12}}{z_{11}z_{22} - z_{12}^2} \quad (12-7-17)$$

The three equations for the z 's in terms of the y 's are

$$z_{11} = \frac{y_{22}}{y_{11}y_{22} - y_{12}^2} \quad (12-7-18)$$

$$z_{22} = \frac{y_{11}}{y_{11}y_{22} - y_{12}^2} \quad (12-7-19)$$

and

$$z_{12} = -\frac{y_{12}}{y_{11}y_{22} - y_{12}^2} \quad (12-7-20)$$

12-8. Poles and Zeros of RC Transfer Functions. Of the parameters discussed in the previous article z_{11} , z_{22} , y_{11} , and y_{22} are either driving-point impedances or admittances of an RC network; and the locations of the poles and zeros of these types of functions have already been discussed. However, z_{12} and y_{12} are transfer functions and nothing has yet been said about the locations of their poles and zeros. Based on an idea suggested by Cauer and developed by Brune, the condition for physical realizability for any transfer function can be found by reducing the problem to a driving-point impedance by the artifice shown in Fig. 12-8-1. For the present purposes, the discussion is limited to RC networks.

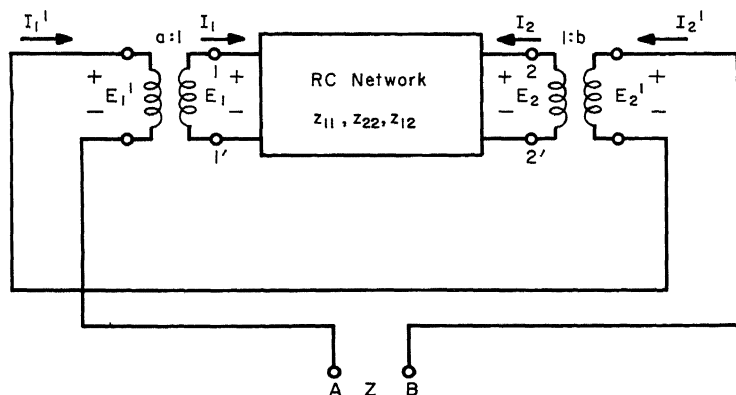


Fig. 12-8-1.

The RC network is described by z_{11} , z_{22} , and z_{12} . A perfect transformer with a turns ratio of $a:1$ is connected to the 1-1' terminals and another perfect transformer with a turns ratio of $b:1$ is connected to the 2-2' terminals. Since the entire circuit is made up of R 's and C 's, the driving-point impedance Z at terminals A-B must be that of an RC network. An expression for the impedance Z is found in the following steps.

The original network is described by equations

$$E_1 = z_{11}I_1 + z_{12}I_2 \quad (12-8-1)$$

$$E_2 = z_{12}I_1 + z_{22}I_2 \quad (12-8-2)$$

These equations can be rewritten in terms of the voltage and currents on the other sides of the transformers as

$$\frac{E'_1}{a} = z_{11}aI'_1 + z_{12}bI'_2 \quad (12-8-3)$$

$$\frac{E'_2}{a} = z_{12}aI'_1 + z_{22}bI'_2 \quad (12-8-4)$$

Because of the manner in which the circuit is connected,

$$I'_1 = I'_2 \quad (12-8-5)$$

and the equations can be rewritten

$$E'_1 = z_{11}a^2I'_1 + z_{12}abI'_1 \quad (12-8-6)$$

$$E'_2 = z_{12}abI'_1 + z_{22}b^2I'_1 \quad (12-8-7)$$

The driving-point impedance is

$$Z = \frac{E'_1 + E'_2}{I'_1} \quad (12-8-8)$$

When Eqs. 12-8-6 and 12-8-7 are substituted, the equation for Z is obtained as

$$Z = \frac{z_{11}a^2I'_1 + 2z_{12}abI'_1 + z_{22}b^2I'_1}{I'_1} \quad (12-8-9)$$

or as

$$Z = z_{11}a^2 + 2z_{12}ab + z_{22}b^2 \quad (12-8-10)$$

The location of the poles and zeros of Z , z_{11} , and z_{22} are known because these are RC driving-point impedances. The location of the poles and zeros of z_{12} can be deduced by the following steps.

Eq. 12-8-10 is solved for $2z_{12}ab$:

$$2z_{12}ab = Z - z_{11}a^2 - z_{22}b^2 \quad (12-8-11)$$

Since Z , z_{11} , and z_{22} all have simple poles on the negative real axis, z_{12} must also have simple poles on the negative real axis.

It can also be determined by examining Eq. 12-8-11, that all the coefficients in the numerator polynomial of z_{12} must be real; however, some of these coefficients may be negative.

As explained in the discussion of Foster's first form, when an RC driving-point impedance is expanded in partial fractions, all the residue terms are positive. For example, if

$$s = -p_1, -p_2, -p_3 \quad (12-8-12)$$

are poles of Z , then Z can be expanded on

$$Z = \frac{K_{p_1}}{s + p_1} + \frac{K_{p_2}}{s + p_2} + \frac{K_{-p_3}}{s + p_3} \quad (12-8-13)$$

where each of the K 's is positive. If s is allowed to approach the value $-p_1$ the first term on the right of Eq. 12-8-13 becomes so large compared with the other two terms that Z behaves in essentially the same manner as the first term. In a similar manner, if $s = -p_1$ is a pole common to Z , z_{11} , z_{22} , and z_{12} , then as s approaches $-p_1$, each of the terms of Eq. 12-8-10 (repeated here for convenience):

$$Z = z_{11}a^2 + 2z_{12}ab + z_{22}b^2 \quad (12-8-14)$$

behaves as

$$\frac{K}{s + p_1} = \frac{K_{11}a^2}{s + p_1} + \frac{K_{12}ab}{s + p_1} + \frac{K_{22}b^2}{s + p_1} \quad (12-8-15)$$

From this equation it can be seen that

$$K = K_{11}a^2 + K_{12}ab + K_{22}b^2 \quad (12-8-16)$$

where K , K_{11} , and K_{22} are positive real numbers. Regardless of the size or signs of a and b , the right side of this equation must be positive. From the theory of positive definite quadratic forms, this leads to

$$K_{11}K_{22} - K_{12}^2 \geq 0 \quad (12-8-17)$$

This equation is referred to as the residue condition.

If K_{12} is zero, the residue condition can be satisfied with K_{11} and K_{22} taking on any positive value including zero. However, if K_{12} is not zero, then neither K_{11} or K_{22} can be zero. From this it can be reasoned that z_{11} can have a pole not appearing in z_{12} or in z_{22} , and that z_{22} can have a pole not appearing in z_{12} or in z_{11} . However, if z_{12} has a pole, this pole must appear in both z_{11} and z_{22} .

It should also be noted that K_{12} is real but there is no restriction on its sign. Therefore K_{12} can be negative.

From this discussion it can be seen that the poles of z_{12} must be simple and on the negative real axis but that the zeros can be anywhere and can be of any multiplicity. When the zeros are complex, they must appear in conjugate pairs.

The final point to be considered, is the behavior of z_{12} as s approaches infinity. Eq. 12-8-10 is again referred to for this discussion, being repeated as

$$Z = z_{11}a^2 + 2z_{12}ab + z_{22}b^2 \quad (12-8-18)$$

As s approaches infinity, each of the driving-point impedances is either a resistance or a short-circuit. For the moment, each driving-point impedance is assumed to be a resistance. Therefore, regardless of the

size or sign of a and b , the right side of Eq. 12-8-18 is positive as $s \rightarrow \infty$. From this a condition somewhat similar to the residue condition results:

$$z_{11}(\infty)z_{22}(\infty) - z_{12}^2(\infty) \geq 0 \quad (12-8-19)$$

If $z_{12}(\infty)$ is zero, $z_{11}(\infty)$ and $z_{22}(\infty)$ can have any values including zero. However, if either $z_{11}(\infty)$ or $z_{22}(\infty)$ is zero, then $z_{12}(\infty)$ must also be zero. Also if $z_{12}(\infty)$ has a value other than zero, then neither $z_{11}(\infty)$ nor $z_{22}(\infty)$ can be zero.

The realizability conditions for y_{11} , y_{22} , and y_{12} can be developed in a similar manner. However, if these parameters are divided by s as y_{11}/s , y_{22}/s , and y_{12}/s , these new functions satisfy exactly the same set of conditions discussed for the z 's.

An example best illustrates the way the conditions on the z 's can be used.

Example 12-8-1. The following z 's are given:

$$z_{11} = \frac{(s+1)(s+3)}{s(s+2)(s+4)} \quad (12-8-20)$$

$$z_{22} = \frac{(s+3)(s+5)}{(s+2)(s+4)} \quad (12-8-21)$$

$$z_{12} = \frac{H(s+1)}{(s+2)(s+4)} \quad (12-8-22)$$

It is desired to determine the largest possible value for the constant H without violating the residue or $s = \infty$ conditions.

The condition

$$z_{11}(\infty)z_{22}(\infty) - z_{12}^2(\infty) \geq 0 \quad (12-8-23)$$

is satisfied with an equality sign as both $z_{11}(\infty)$ and $z_{12}(\infty)$ are zero.

To continue the example, the z 's are expanded as

$$z_{11} = \frac{0.375}{s} + 0 + \frac{0.250}{s+2} + \frac{0.375}{s+4} \quad (12-8-24)$$

$$z_{22} = 0 + 1 + \frac{1.5}{s+2} + \frac{0.5}{s+4} \quad (12-8-25)$$

$$z_{12} = 0 + 0 - \frac{0.5H}{s+2} + \frac{1.5H}{s+4} \quad (12-8-26)$$

The residue condition $K_{11}K_{22} - K_{12}^2 \geq 0$ is applied at the pole at $s = -2$ as

$$(0.250)(1.5) - (0.5H)^2 \geq 0 \quad (12-8-27)$$

from which

$$H \leq 1.225 \quad (12-8-28)$$

This is repeated for the pole at $s = -4$ as

$$(0.375)(0.5) - (1.5H)^2 \geq 0 \quad (12-8-29)$$

from which

$$H \leq 0.288 \quad (12-8-30)$$

Therefore, the maximum possible value for H is 0.288 in order to synthesize the three z 's as an RC network.

In this book, no synthesis procedures are attempted in which all three z 's (or y 's) are specified. Such procedures almost invariably involve perfect transformers as circuit elements. The use of the perfect transformer can be avoided in many situations by using synthesis procedure in which two of the three z 's (or y 's) are specified. This is explored in the following articles.

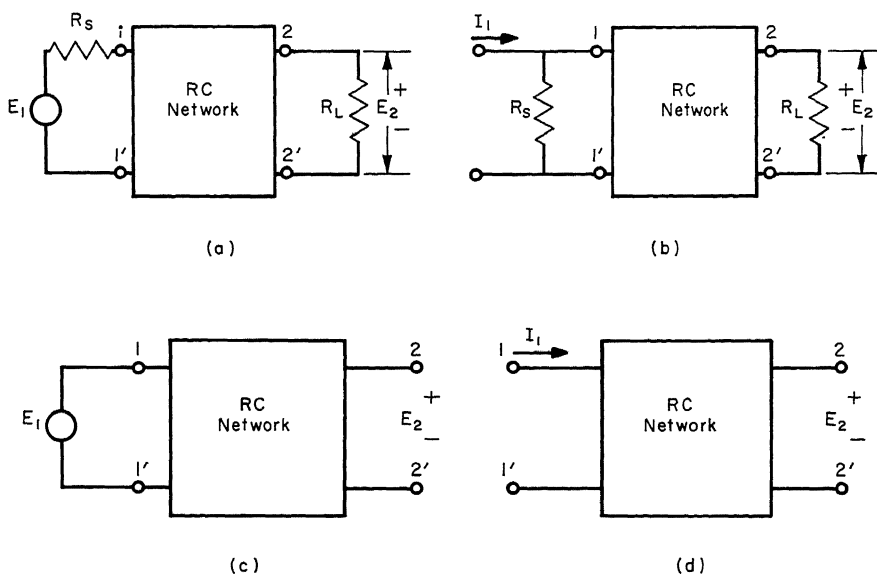


Fig. 12-9-1.

12-9. Introductions to Synthesis Procedures for Transfer Functions. In control systems, the RC network is usually inserted in the system. Such a network is shown in Fig. 12-9-1a and the Norton theorem equivalent in Fig. 12-9-1b. If the network is driven by a triode, R_s is finite. If it is driven by a pentode, the circuit of Fig. 12-9-2b is preferred for representing the circuit, and in many cases R_s can be assumed to be infinite when compared with the impedance at the 1-1' terminals. In many cases, if the network is driven by a cathode follower or by some other low-impedance device, R_s can be assumed to be zero when compared with the impedance at the 1-1' terminals.

The output terminals may be connected to a finite load R_L or to the grid of the next tube or some other high-impedance device in which case R_L can be assumed to be infinite.

Since the network itself contains resistors, the R_s and R_L can be considered to be part of the network. For example, if R_s is finite, the network of Fig. 12-9-1c can be used but the RC network must be synthesized so that it begins with a resistor. All or part of this resistor is identified as R_s . Similarly if R_L is finite, the RC network must be synthesized so that it terminates in a resistor, and all or part of this resistor is identified as R_L . Therefore the network of Fig. 12-9-1c with the proper synthesis procedure can be used to obtain the network of Fig. 12-9-1a; likewise, the network of Fig. 12-9-1d can be used to obtain the network of Fig. 12-9-1b.

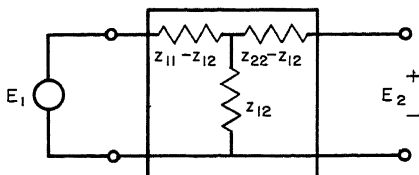


Fig. 12-9-2.

For the present purposes, the network of Fig. 12-9-1c is considered. The RC portion of the circuit is replaced by an equivalent T as shown in Fig. 12-9-2. The voltage E_2 in terms of E_1 is

$$E_2 = \frac{z_{12}}{z_{11}} E_1 \quad (12-9-1)$$

the transfer function is

$$\frac{E_2}{E_1} = \frac{z_{12}}{z_{11}} \quad (12-9-2)$$

From Eqs. 12-7-18 and 12-7-20, the transfer function can be re-written

$$\frac{E_2}{E_1} = -\frac{y_{12}(y_{11}y_{22} - y_{12}^2)}{y_{22}(y_{11}y_{22} - y_{12}^2)} = -\frac{y_{12}}{y_{22}} \quad (12-9-3)$$

The minus sign is the result of the direction chosen for currents and can be ignored for the present purposes.

The conditions necessary for E_2/E_1 to be realized with an RC network can be determined from either Eq. 12-9-2 or Eq. 12-9-3. For this discussion, Eq. 12-9-2 is used. The zeros of E_2/E_1 are either the zeros of z_{12} or the poles of z_{11} not in z_{12} . The poles of E_2/E_1 are the zeros of z_{11} not in z_{12} . The poles in z_{12} do not create poles in E_2/E_1 because each of these poles also appears in z_{11} .

The poles of E_2/E_1 are the zeros of z_{11} which occur on the negative real axis. Since the lowest critical frequency of z_{11} is a pole, z_{11} cannot have a zero at the origin and hence E_2/E_1 does not have a pole at the origin. If z_{11} has a zero at infinity, by Eq. 12-8-19 z_{12} must also have a zero; therefore, E_2/E_1 cannot have a pole at infinity.

12-10. Transfer Functions with Zeros on the Negative Real Axis.

If the zeros of E_2/E_1 are on the negative real axis, a ladder network can be used to realize the transfer function. A ladder network is shown in Fig.

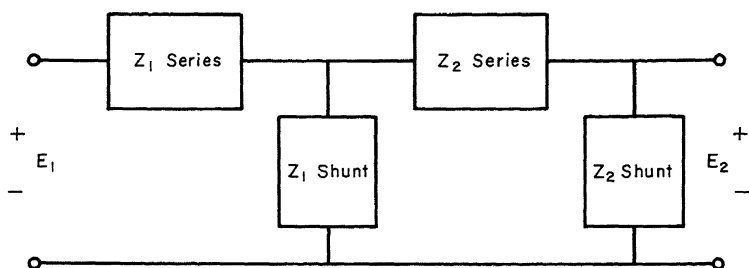


Fig. 12-10-1.

12-10-1 with two series and two shunt impedances. The zeros in transmission are created either by a series impedance becoming an open circuit or by a shunt impedance becoming a short circuit. The necessary additional comments are included in the discussion of the examples.

Example 12-10-1. The following transfer function is given

$$\frac{E_2}{E_1} = \frac{(s+1)(s+8)}{(s+2)(s+4)} \quad (12-10-1)$$

The network to realize this E_2/E_1 is first to be synthesized without regard to loading. It is decided to use the identification of Eq. 12-9-3:

$$\frac{E_2}{E_1} = \frac{y_{12}}{y_{22}} = \frac{(s+1)(s+8)/q(s)}{(s+2)(s+4)/q(s)} \quad (12-10-2)$$

from which

$$y_{12} = \frac{(s+1)(s+8)}{q(s)} \quad y_{22} = \frac{(s+2)(s+4)}{q(s)} \quad (12-10-3)$$

The polynomial $q(s)$ must be chosen so that y_{22} is physically realizable as an RC driving-point admittance. This means that $q(s)$ must contain a factor $s+a$ where a is between 2 and 4. The polynomial $q(s)$ can also contain a second factor $s+\beta$

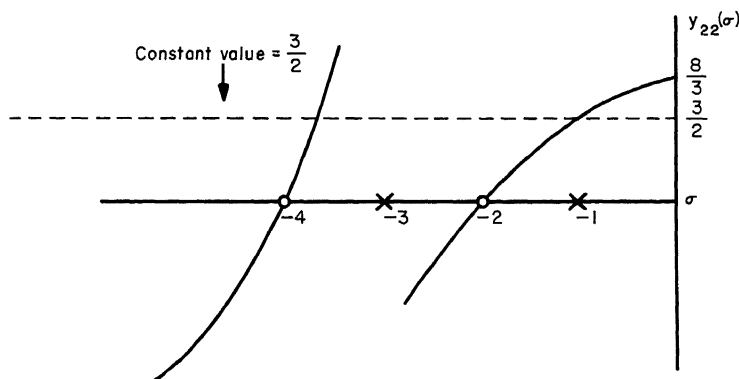


Fig. 12-10-2.

where β is any number larger than 4. To make y_{22} as simple as possible, the $s + a$ factor is used by itself. The constant a is arbitrarily chosen as 3.

$$y_{12} = \frac{(s+1)(s+8)}{s+3} \quad y_{22} = \frac{(s+2)(s+4)}{s+3} \quad (12-10-4)$$

The procedure to be followed is to synthesize y_{22} in such a way that the desired transmission zeros of y_{12} are realized. As a help in explaining this procedure, a plot of y_{22} versus σ is shown in Fig. 12-10-2. The zero at -2 in y_{22} can be moved to -1 by subtracting the proper constant value from $y_{22}(s)$. This constant value is found as

$$y_{22}(s) \Big|_{s=-1} = \frac{(-1+2)(-1+4)}{-1+3} = \frac{3}{2} \quad (12-10-5)$$

This value is subtracted from y_{22} leaving y_A

$$y_A = \frac{(s+2)(s+4)}{s+3} - \frac{3}{2} = \frac{(s+1)\left(s+\frac{7}{2}\right)}{s+3} \quad (12-10-6)$$

The above steps have created a zero at $s = -1$ in y_A and with y_A "flipped over," the result has a pole at $s = -1$

$$\frac{1}{y_A} = \frac{s+3}{(s+1)\left(s+\frac{7}{2}\right)} \quad (12-10-7)$$

This pole is removed by calculating the residue in the pole at -1 as

$$k_{-1} = \frac{-1+3}{-1+\frac{7}{2}} = \frac{4}{5} \quad (12-10-8)$$

The resulting function is labeled $1/y_B$

$$\frac{1}{y_B} = \frac{s+3}{(s+1)\left(s+\frac{7}{2}\right)} - \frac{\frac{4}{5}}{s+1} = \frac{\frac{1}{5}}{s+\frac{7}{2}} \quad (12-10-9)$$

from which y_B is obtained:

$$y_B = 5\left(s+\frac{7}{2}\right) \quad (12-10-10)$$

A plot of y_B versus σ is shown in Fig. 12-10-3. The zero at $-\frac{7}{2}$ is moved to -8

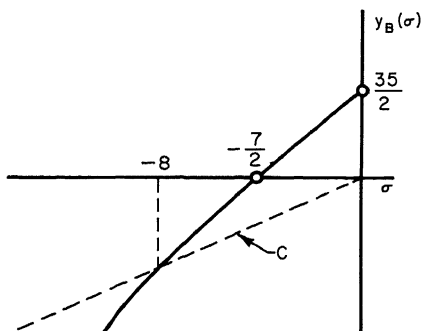


Fig. 12-10-3.

by subtracting Cs from y_B . C is evaluated as

$$Cs|_{s=-8} = y_B|_{s=-8} = 5(-8 + \frac{7}{2}) = -8C \quad (12-10-11)$$

$$C = \frac{45}{16} \quad (12-10-12)$$

This Cs is subtracted from y_B leaving y_C

$$y_C = 5(s + \frac{7}{2}) - \frac{45}{16}s = \frac{35}{16}(s + 8) \quad (12-10-13)$$

The following expansion has been performed on y_{22} .

$$y_{22} = \frac{3}{2} + \frac{1}{\frac{4/5}{s+1} + \frac{1}{\frac{45}{16}s + \frac{35}{16}(s+8)}} \quad (12-10-14)$$

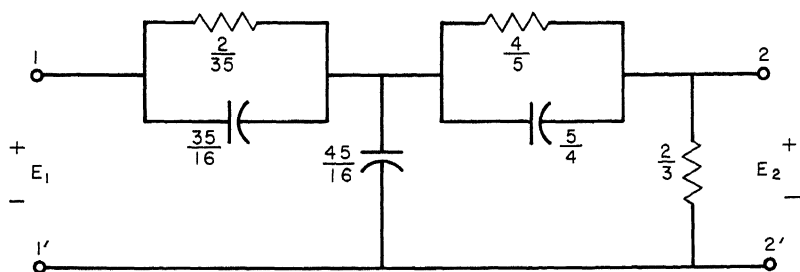


Fig. 12-10-4.

The network realizing y_{22} is shown in Fig. 12-10-4. The admittance y_{22} is found at the 2-2' terminals with the 1-1' terminals shorted. A comparison of Fig. 12-10-4 with Eq. 12-10-14 indicates that this is so. However, the circuit is judiciously arranged so that the proper zeros of y_{12} also are realized. When E_1 is applied to the 1-1' terminals, the $R = 2/35$, $C = 35/16$ combination creates a zero in transmission at $s = -8$ and the $R = 4/5$, $C = 5/4$ combination creates a zero at $s = -1$.

In order for a zero in transmission to be realized by a pole removal as in Eq. 12-10-9, all of the pole must be removed. If only a part of the pole is removed, the resulting network acts as a voltage divider. Fig. 12-10-5 shows a simple example.

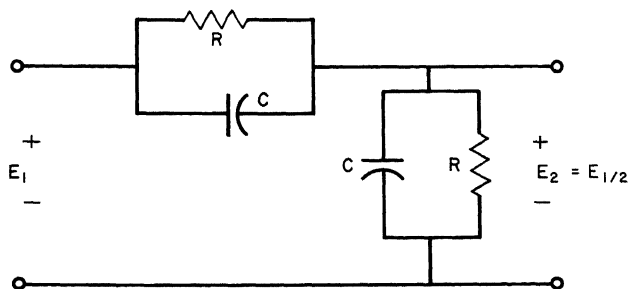


Fig. 12-10-5.

The RC combination in the series portion of the ladder does not create the desired zero in transmission because the remainder of the network also contains the same pole. Similarly when a shunt arm is removed to create a zero in transmission, all the pole in the corresponding admittance function must be removed. Otherwise, the current entering this portion of the circuit will divide between two short circuits and will affect the output.

The transfer admittance y_{12} is realized only within a constant multiplier. This results from the fact that y_{12} is not realized as such but only the zeros of y_{12} are realized. Therefore to be correct, y_{12} of Eq. 12-10-4 should be written

$$y_{12} = \frac{K(s+1)(s+8)}{s+3} \quad (12-10-15)$$

The constant K can be evaluated from the network of Fig. 12-10-4. This is most easily done by letting s approach zero. From Eq. 12-10-15, y_{12} becomes

$$y_{12} \Big|_{s \rightarrow 0} = K \frac{8}{3} \quad (12-10-16)$$

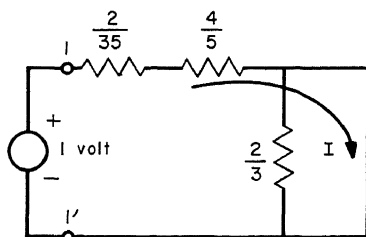


Fig. 12-10-6.

and the network degenerates to that shown in Fig. 12-10-6. The transfer admittance is determined from this circuit as

$$y_{12} = \frac{7}{6} \quad (12-10-17)$$

When Eqs. 12-10-16 and 12-10-17 are compared, K is determined as

$$K = \frac{7}{16} \quad (12-10-18)$$

Example 12-10-2. The same transfer function

$$\frac{E_2}{E_1} = \frac{K(s+1)(s+8)}{(s+2)(s+4)} \quad (12-10-19)$$

and the same y_{12} and y_{22}

$$y_{12} = \frac{K(s+1)(s+8)}{s+3} \quad y_{22} = \frac{(s+2)(s+4)}{s+3} \quad (12-10-20)$$

of Example 12-10-1 are realized by a different development. The zero in transmission at $s = -8$ is realized first. The zero at $s = -4$ in y_{22} is moved to $s = -8$ by removing a capacitor. It is suggested that the reader draw the appropriate sketches

of the y 's and $1/y$'s to visualize each of the following steps. The value of C is found by setting up the equation

$$Cs \Big|_{s=-8} = -8C = y_{22} \Big|_{s=-8} = -\frac{24}{5} \quad (12-10-21)$$

and solving for C as

$$C = \frac{3}{5} \quad (12-10-22)$$

This Cs is subtracted from y_{22} leaving y_A

$$y_A = \frac{(s+2)(s+4)}{s+3} - \frac{3}{5}s = \frac{\left(\frac{2}{5}s + 1\right)(s+8)}{s+3} \quad (12-10-23)$$

The admittance y_A is "flipped-over"

$$\frac{1}{y_A} = \frac{s+3}{\left(\frac{2}{5}s + 1\right)(s+8)} \quad (12-10-24)$$

The residue in the pole at $s = -8$ is calculated as

$$k_{-8} = \frac{-5}{-\frac{16}{5} + 1} = \frac{25}{11} \quad (12-10-25)$$

The pole is removed and the remainder function is labeled $1/y_B$

$$\frac{1}{y_B} = \frac{s+3}{\left(\frac{2}{5}s + 1\right)(s+8)} - \frac{25/11}{s+8} = \frac{1}{\frac{22}{5}\left(s + \frac{5}{2}\right)} \quad (12-10-26)$$

The zero at $s = -5/2$ in

$$y_B = \frac{22}{5}\left(s + \frac{5}{2}\right) \quad (12-10-27)$$

is moved to $s = -1$ by removing a resistor. The value of R is found by setting up the equation

$$y_B \Big|_{s=-1} = \frac{22}{5}\left(-1 + \frac{5}{2}\right) = \frac{33}{5} \quad (12-10-28)$$

After this R is removed from y_B , the remainder function is to be called y_C :

$$y_C = \frac{22}{5}\left(s + \frac{5}{2}\right) - \frac{33}{5} = \frac{22}{5}(s+1) \quad (12-10-29)$$

The following expansion has been performed on y_{22}

$$y_{22} = \frac{3}{5}s + \frac{1}{\frac{25/11}{s+8} + \frac{\frac{33}{5} + \frac{22}{5}(s+1)}{1}} \quad (12-10-30)$$

Again the circuit is judiciously arranged to yield the proper y_{22} with the 1-1' terminals shorted to realize the zeros of y_{12} . The resulting circuit is shown in Fig. 12-10-7.

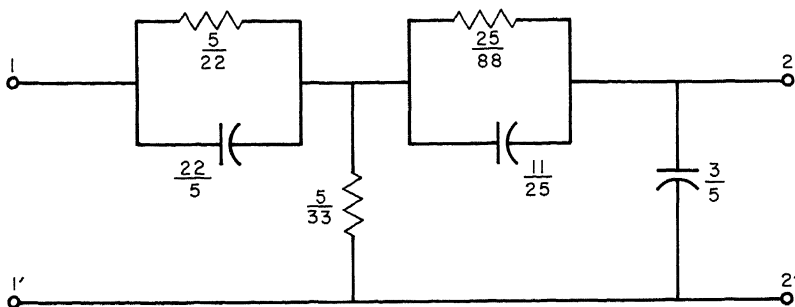


Fig. 12-10-7.

The constant K associated with y_{12} is determined by letting s approach infinity. The y_{12} from Eq. 12-10-20 becomes

$$y_{12} \big|_{s \rightarrow \infty} = Ks \quad (12-10-31)$$

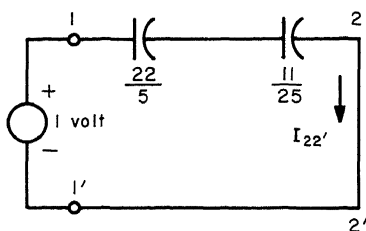


Fig. 12-10-8.

As $s \rightarrow \infty$, the circuit of Fig. 12-10-7 becomes that of Fig. 12-10-8 as far as determining K is concerned. The current I'_{22} is

$$y_{12} \big|_{s \rightarrow \infty} = I'_{22} = 0.4s \quad (12-10-32)$$

From a comparison of Eq. 12-10-31 and 12-10-32, K is determined as

$$K = 0.4 \quad (12-10-33)$$

Example 12-10-3. The same transfer function

$$\frac{E_2}{E_1} = \frac{K(s+1)(s+8)}{(s+2)(s+4)} \quad (12-10-34)$$

is realized except that the development is carried out in terms of z 's.

$$\frac{E_2}{E_1} = \frac{K(s+1)(s+8)/q(s)}{(s+2)(s+4)/q(s)} = \frac{z_{12}}{z_{11}} \quad (12-10-35)$$

The following identifications are made

$$z_{11} = \frac{(s+2)(s+4)}{q(s)} ; \quad z_{12} = \frac{K(s+1)(s+8)}{q(s)} \quad (12-10-36)$$

The polynomial $q(s)$ must be chosen so that z_{11} is an RC driving-point impedance. To simplify z_{12} , one factor of $q(s)$ is chosen as $s + 1$ and the other factor is arbitrarily chosen as $s + 3$. The z 's to be realized are

$$z_{11} = \frac{(s+2)(s+4)}{(s+1)(s+3)} \quad ; \quad z_{12} = \frac{K(s+8)}{s+3} \quad (12-10-37)$$

The pole in z_{11} at $s = -1$ is removed first. The residue in this pole is determined as $k_{-1} = 3/2$ and the pole is removed

$$z_A = \frac{(s+2)(s+4)}{(s+1)(s+3)} - \frac{3/2}{s+1} = \frac{s+7/2}{s+3} \quad (12-10-38)$$

A convenient way to proceed is to move the zero at $s = -7/2$ to $s = -8$. The z_A is evaluated at $s = -8$ as $9/10$ and this amount is removed as

$$z = \frac{s+7/2}{s+3} - \frac{9}{10} = \frac{(1/10)(s+8)}{s+3} \quad (12-10-39)$$

The z_B is "flipped over" and $1/z_B s$ is expanded and the result multiplied by s to yield

$$\frac{1}{z_B} = \frac{15}{4} + \frac{(25/4)s}{s+8} \quad (12-10-40)$$

The following expansion has been performed on z_{11}

$$z_{11} = \frac{3/2}{s+1} + \frac{9}{10} + \frac{1}{\frac{(25/4)s}{s+8} + \frac{15}{4}} \quad (12-10-41)$$

The resulting circuit is rearranged slightly and is shown in Fig. 12-10-9.

The gain K can be calculated to be

$$K = 0.1 \quad (12-10-42)$$

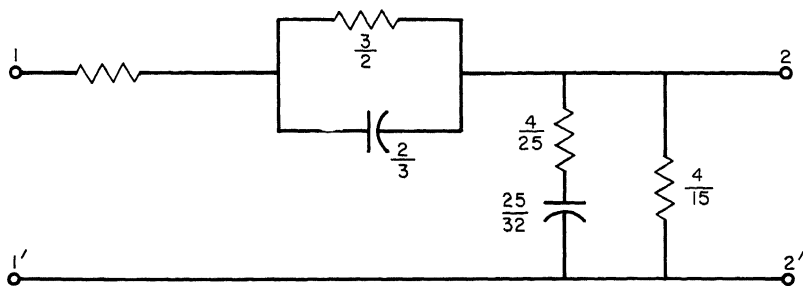


Fig. 12-10-9.

DISCUSSION OF THE PREVIOUS EXAMPLES. In the previous three examples, the y 's and the z 's are developed without regard to loading on either end of the network. The circuit for the first example is shown in Fig. 12-10-4. This circuit can be readily adapted to loading on the receiving end by moving the $2/3$ ohm resistor external to the $2-2'$ terminals and by scaling the circuit impedances so that the resistor equals the desired load resistance.

The circuit for the second example is shown in Fig. 12-10-7. This circuit cannot be adapted to loading at either end.

The circuit for the third example is shown in Fig. 12-10-9. The $4/15$ ohm can be moved external to the $2-2'$ terminals for loading at the receiving end and the $9/10$ ohm resistor can be moved external to the $1-1'$ terminals for loading at the sending end. The circuit impedance can be scaled so that the loading impedance demands at both ends of the network are met. If for example, in doing this, the $9/10$ ohm resistor is scaled to a larger value than demanded by the loading, a resistor equal to the difference between the two values is left internal in the RC network.

When loading is demanded at one or both ends of the network, the designer must anticipate the appropriate steps in the synthesis procedure so as to realize the loading. This point of view is discussed in the next example.

Example 12-10-4. The same transfer function

$$\frac{E_2}{E_1} = \frac{K(s+1)(s+8)}{(s+2)(s+4)} \quad (12-10-43)$$

is realized in terms of y_{12} and y_{22} but it is desirable to have loading at both ends of the network.

For the loading at the receiving end, the synthesis procedure can begin as in Example 12-10-1. However, in Example 12-10-1, both y_{12} and y_{22} have poles at infinity and this cannot possibly lead to loading at the sending end. Therefore, another pole is chosen outside the zero at -4 to yield the y 's as

$$y_{22} = \frac{(s+2)(s+4)}{(s+3)(s+6)} \quad ; \quad y_{12} = \frac{K(s+1)(s+8)}{(s+3)(s+6)} \quad (12-10-44)$$

The zero at $s = -2$ in y_{22} is moved to $s = -1$. The y_{22} is evaluated at $s = -1$ as $3/10$ and this amount is removed as

$$y_A = \frac{(s+2)(s+4)}{(s+3)(s+6)} - \frac{3}{10} = \frac{(s+1)\left(\frac{7}{10}s + \frac{13}{5}\right)}{(s+3)(s+6)} \quad (12-10-45)$$

The y_A is "flipped over" and the pole at $s = -1$ is removed. The residue in this pole is determined as $k_{-1} = 100/19$ and the pole is removed

$$\frac{1}{y_B} = \frac{(s+3)(s+6)}{(s+1)\left(\frac{7}{10}s + \frac{13}{5}\right)} - \frac{100/19}{s+1} = \frac{s+82/19}{\frac{7}{10}s + \frac{13}{5}} \quad (12-10-46)$$

A convenient way to proceed is to move the zero at $s = -82/19$ to $s = -8$. The $1/y_B$ is evaluated at $s = -8$ as $70/57$ and this amount is removed as

$$\frac{1}{y_C} = \frac{s+82/19}{\frac{7}{10}s + \frac{13}{5}} - \frac{70}{57} = \frac{(8/57)(s+8)}{\frac{7}{10}s + \frac{13}{5}} \quad (12-10-47)$$

The impedance is “flipped over” and y_C/s is expanded and the result multiplied by s to yield.

$$y_C = \frac{171s}{64(s+8)} + \frac{741}{320} \quad (12-10-48)$$

The following expansion has been performed on y_{22} :

$$y_{22} = \frac{3}{10} + \frac{1}{\frac{100/19}{s+1} + \frac{70}{57} + \frac{1}{\frac{171s}{64(s+8)} + \frac{741/320}} \quad (12-10-49)$$

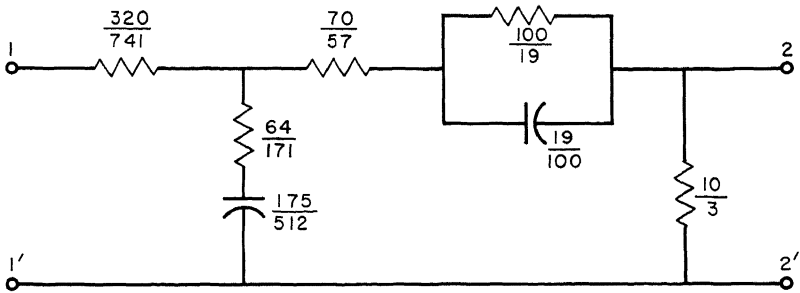


Fig. 12-10-10.

The circuit is judiciously arranged as shown in Fig. 12-10-10. The gain K can be calculated to be

$$K = 0.325 \quad (12-10-50)$$

The desired transfer function has been realized in such a manner that loading resistors are available at each end of the network.

Example 12-10-5. Since the single ladder network achieves zeros in transmission by the series impedance opening or the shunt impedance shorting, this type of structure can produce zeros only on the negative real axis. If the zeros are complex conjugate pairs, the transfer function can be realized on the admittance basis by a parallel ladder development as demonstrated by the example

$$\frac{E_2}{E_1} = \frac{K(s^2 + 2)}{(s+2)(s+4)} = \frac{y_{12}}{y_{22}} \quad (12-10-51)$$

The following identifications are made

$$y_{12} = \frac{K(s^2 + 2)}{q(s)} ; \quad y_{22} = \frac{(s + 2)(s + 4)}{q(s)} \quad (12-10-52)$$

and $q(s)$ is arbitrarily chosen as $s + 3$:

$$y_{12} = \frac{K(s^2 + 2)}{s + 3} ; \quad y_{22} = \frac{(s + 2)(s + 4)}{s + 3} \quad (12-10-53)$$

The y_{12} and y_{22} are divided into two parts as

$$y_{12} = \frac{K(s^2 + 2)}{s + 3} = \frac{aK_a s^2}{s + 3} + \frac{\beta K_b 2}{s + 3} \quad (12-10-54)$$

and

$$y_{22} = \frac{(s + 2)(s + 4)}{s + 3} = \frac{a(s + 2)(s + 4)}{s + 3} + \frac{\beta(s + 2)(s + 4)}{s + 3} \quad (12-10-55)$$

The a and β networks are developed separately and are connected in parallel to achieve the desired y 's. The K_a and K_b are the gains that result from the synthesis procedures on the two separate networks. The a and β factors are gain adjustment factors that are necessary to make the total gain aK_a and βK_b equal as

$$aK_a = \beta K_b \quad (12-10-56)$$

The other equation evolving a and β is a result of Eq. 12-10-55 as

$$a + \beta = 1 \quad (12-10-57)$$

The a network is designed from

$$(y_{12})_a = \frac{K_a s^2}{s + 3} ; \quad (y_{22})_a = \frac{(s + 2)(s + 4)}{s + 3} \quad (12-10-58)$$

The $(y_{22})_a$ is developed so that two zeros in transmission at the origin are realized. This can be done by developing $(y_{22})_a$ in Cauer's second form as

$$(y_{22})_a = \frac{8}{3} + \frac{1}{\frac{9}{10s} + \frac{1}{\frac{100}{3} + \frac{1}{\frac{1}{10s}}}} \quad (12-10-59)$$

which leads to the network of Fig. 12-10-11.

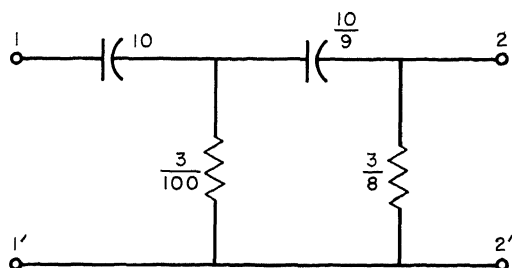


Fig. 12-10-11.

The β network is designed from

$$(y_{12})_{\beta} = \frac{K_b 2}{(s+3)} ; \quad (y_{22})_{\beta} = \frac{(s+2)(s+4)}{(s+3)} \quad (12-10-60)$$

The $(y_{22})_{\beta}$ is developed so that one zero in transmission at infinity is realized. This can be done by developing $(y_{22})_{\beta}$ in Cauer's first form as

$$(y_{22})_{\beta} = s + \frac{1}{\frac{1}{3} + \frac{1}{9s + \frac{1}{\frac{1}{24}}}} \quad (12-10-61)$$

which leads to the network of Fig. 12-10-12.

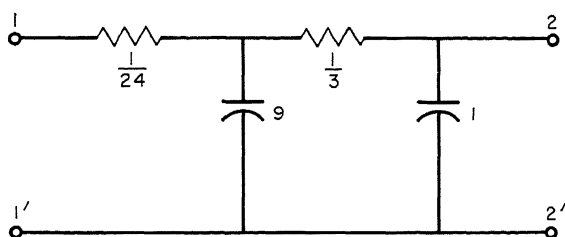


Fig. 12-10-12.

The gain K_a can be determined for the network of Fig. 12-10-11 as

$$K_a = 1 \quad (12-10-62)$$

The gain K_b can be determined for the network of Fig. 12-10-12 as

$$K_b = 4$$

If the two networks were connected in parallel, the result would be

$$y_{12} = \frac{K_a s^2 + K_b^2}{s+3} = \frac{s^2 + 8}{s+3} \quad (12-10-63)$$

and the desired y_{12} would not be realized. However, the admittance level of each network can be adjusted by the a and β factors, which are determined from

$$aK_a = \beta K_b \quad (12-10-64)$$

and from

$$a + \beta = 1 \quad (12-10-65)$$

upon substituting the known values for K_a and K_b , these equations can be solved for a and β , yielding

$$a = \frac{4}{5}, \quad \beta = \frac{1}{5} \quad (12-10-66)$$

The admittance of a resistor and capacitor are respectively

$$\frac{1}{R} \quad \text{and} \quad Cs \quad (12-10-67)$$

To change the admittance level of the α network by a factor of $4/5$, the R 's are multiplied by $5/4$ and the C 's by $4/5$.

To change the admittance level of the β network by I factor of $1/5$, the R 's are multiplied by 5 and the C 's by $1/5$.

After the admittance levels are adjusted, the two networks are connected in parallel as shown in Fig. 12-10-13.

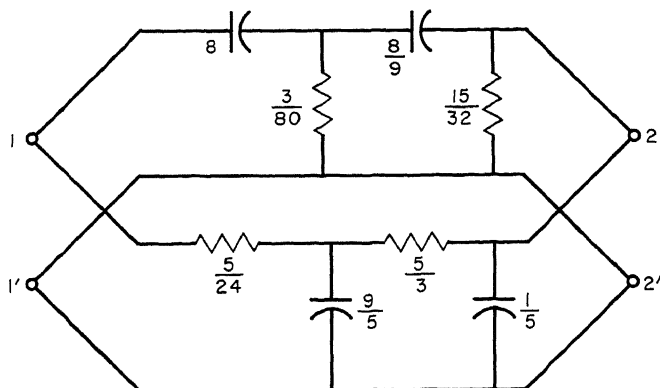


Fig. 12-10-13.

With the 1-1' terminals shorted, y_{22} is the sum of the admittance of the two separate networks. The current that exists in the short between the 1-1' terminals is the sum of the currents from the individual networks; therefore, y_{12} is the sum of the transfer admittance of the individual networks. No analogous method is possible on the impedance basis.

The gain K for the entire circuit is

$$K = aK_a = \beta K_b = \frac{4}{5} \quad (12-10-68)$$

12-11. Concluding Comments. This chapter is intended as an introduction to network synthesis. Only enough concepts are explained here to provide the background necessary for the next chapter. For a more complete coverage of network synthesis, the reader should consult one of the books devoted exclusively to this subject.

PROBLEMS

12-1. Determine which of the following functions can be synthesized as an LC driving-point impedances:

(a) $\frac{s(s^2 + 4)(s^2 + 16)}{(s^2 + 1)(s^2 + 9)}$

(b) $\frac{(s^2 + 1)(s^2 + 9)}{(s^2 + 4)(s^2 + 16)}$

- (c) $\frac{s(s^2 + 1)(s^2 + 3)}{(s^2 + 2)(s^2 + 4)}$ (e) $\frac{s^5 + 40s^3 + 300s}{s^4 + 25s^2 + 100}$
- (d) $\frac{(s^2 + 1)(s^2 + 3)}{s(s^2 + 2)(s^2 + 4)}$ (f) $\frac{s^5 + 30s^2 + 40s}{s^4 + 20s^2 + 90}$

12-2. Some of the following functions can be synthesized as *RC* and others as *RL* driving-point impedances and still others by neither type of circuit. Classify each part.

- (a) $\frac{(s + 2)(s + 4)}{(s + 1)(s + 3)}$ (d) $\frac{(s^2 + 1)(s^2 + 3)}{(s^2 + 2)(s^2 + 4)}$
- (b) $\frac{s + 2}{(s + 1)(s + 3)}$ (e) $\frac{(s + 2)(s + 3)}{(s + 1)(s + 4)}$
- (c) $\frac{s(s + 2)(s + 4)}{(s + 1)(s + 3)}$ (f) $\frac{(s + 1)(s + 3)}{s(s + 2)(s + 4)}$

12-3. Each of the following functions is to be such that it can be synthesized as an *LC*, *RC* or *RL* network. Find the range of a so that they can be so synthesized and identify the type of function for each part.

- (a) $\frac{s(s^2 + a)(s^2 + 25)}{(s^2 + 1)(s^2 + 16)}$ (d) $\frac{(s + 2)(s + 6)}{(s + 1)(s + a)}$
- (b) $\frac{s + 4}{(s + a)(s + 5)}$ (e) $\frac{s(s + 3)}{(s + 2)(s + a)}$
- (c) $\frac{(s^2 + 1)(s^2 + 9)}{s(s^2 + 6)(s^2 + a)}$ (f) $\frac{(s + 1)(s + 8)}{(s + 0.5)(s + a)}$

12-4. Given the function

$$Z(s) = \frac{s(s^2 + 3^2)}{(s^2 + 2^2)(s^2 + 4^2)}$$

find the network for the second Foster form and the first Cauer form, labeling the L 's and C 's.

12-5. Given the function

$$Z(s) = \frac{(s^2 + 2^2)(s^2 + 4^2)}{s(s^2 + 3^2)}$$

find the network for the second Foster form and the first Cauer form, labeling the L 's and C 's.

12-6. Given a Hurwitz polynomial

$$s^5 + s^4 + 6s^3 + 4s^2 + 8s + 3$$

obtain an impedance function by forming the ratio of the even part to the odd part of this polynomial. Find the network for the two Foster and two Cauer forms, labeling the L 's and C 's.

12-7. Repeat Prob. 12-6 except form the impedance function as the ratio of the odd part to the even part.

12-8. Given a Hurwitz polynomial

$$s^6 + 26s^5 + 299s^4 + 1812s^3 + 6086s^2 + 10,780s + 7696$$

obtain an impedance function by forming the ratio of the even part to the odd part of this polynomial. Find the network for the two Cauer forms, labeling the L 's and C 's.

12-9. Repeat Prob. 12-8 except form the impedance function as the ratio of the odd part to the even part.

12-10. Repeat Prob. 12-9 for the polynomial

$$s^6 + 11s^5 + 47s^4 + 101s^3 + 116s^2 + 68s + 16$$

12-11. Synthesize

$$Z(s) = \frac{2(s+2)(s+6)}{s(s+4)(s+8)}$$

as an RC network by Foster's first and second forms.

12-12. Synthesize the function given in Prob. 12-11 as an RC network by Cauer's first and second forms.

12-13. Repeat Prob. 12-11 and 12-12 except the function as given is to be synthesized as an RL admittance.

12-14. Synthesize

$$Z(s) = \frac{s(s+6)(s+10)}{(s+2)(s+8)}$$

as an RL network by using Foster's first and second forms.

12-15. Synthesize the function given in Prob. 12-14 as an RL network by Cauer's first and second forms.

12-16. Repeat Prob. 12-14 and 12-15 except that the function as given is to be synthesized as an RC admittance.

12-17. Synthesize

$$Z(s) = \frac{s+6}{(s+3)(s+10)}$$

as an RC network into all four canonic forms.

12-18. Given the impedance function of Prob. 12-17, calculate the residue in the pole at $s = -3$. Instead of removing all of the pole at $s = -3$, the residue is divided by 2, and only this much of the pole is removed. The remainder function is "flipped over" and divided by s , so as to expand on the admittance basis. The residue in the finite pole is calculated. The residue is divided by 2, and only this much of the finite pole is removed. The remainder function is expanded by Cauer's second form to complete the network.

12-19. Given the impedance function of Prob. 12-17, for $Y(s)$ divided by s calculate the residue in the pole at $s = -6$ but divide the residue by 2, removing only this much of the pole. The remainder function is "flipped over" and the residue in the larger of the two poles is calculated. This residue is divided by 2, and only this much of the pole is removed. The remainder function is expanded by Foster's second form to complete the network.

12-20. The following z 's are given

$$z_{11} = \frac{(s+2)(s+8)}{(s+1)(s+4)} ; \quad z_{22} = \frac{s+3}{(s+1)(s+4)} ; \quad z_{12} = \frac{Hs(s+2)}{(s+1)(s+4)}$$

Determine whether these three z 's form a consistent set that can be used to describe an RC two-terminal-pair network. If so, find the maximum value for H .

12-21. Repeat Prob. 12-20 for the following z 's:

$$z_{11} = \frac{(s+2)(s+8)}{(s+1)(s+4)} ; \quad z_{22} = \frac{s+3}{(s+1)(s+4)} ; \quad z_{12} = \frac{Hs}{(s+1)(s+4)}$$

12-22. Repeat Prob. 12-20 for the following z 's:

$$z_{11} = \frac{(s+2)(s+8)}{(s+1)(s+4)} ; \quad z_{22} = \frac{(s+3)(s+6)}{(s+1)(s+4)} ; \quad z_{12} = \frac{Hs(s+2)}{(s+1)(s+4)}$$

12-23. In Ex. 12-10-1, check the gain constant $K = 7/16$ by letting s go to infinity.

12-24. For the network of Fig. 12-10-4 in Ex. 12-10-1, write the necessary equations and solve for the transfer function E_2/E_1 . From the result check $K = 7/16$.

12-25. In Ex. 12-10-2, check the gain constant $K = 0.4$ by letting s go to zero.

12-26. For the network of Fig. 12-10-7 in Ex. 12-10-2, write the necessary equations and solve for the transfer function E_2/E_1 . From the results check $K = 0.4$.

12-27. In Ex. 12-10-3, check the gain constant $K = 0.1$ by any convenient method.

12-28. In Ex. 12-10-4 check the gain constant $K = 0.325$ by any convenient method.

12-29. Given the transfer function

$$\frac{E_2}{E_1} = \frac{Ks(s+1)}{(s+2)(s+4)}$$

The network to realize this E_2/E_1 is to be synthesized without regard to loading. The ratio E_2/E_1 is to be identified as y_{12}/y_{22} . After the network is developed, determine K .

12-30. Repeat Prob. 12-29 except identify the ratio E_2/E_1 as z_{12}/z_{11} .

12-31. Repeat Prob. 12-29 except that loading is required at both the sending and the receiving ends of this network.

12-32. Repeat Prob. 12-29 except for

$$\frac{E_2}{E_1} = \frac{K(s+6)(s+8)}{(s+2)(s+4)}$$

12-33. Develop a network that realizes the E_2/E_1 of Prob. 12-32 except that loading is required at both ends of the network. Identify E_2/E_1 as z_{12}/z_{11} .

12-34. Repeat Prob. 12-32 except that loading is required at both ends of the network.

12-35. Given the transfer function

$$\frac{E_2}{E_1} = \frac{K(s^2+4)}{(s+1)(s+6)}$$

find the network that yields this transfer function.

12-36. Given the transfer function

$$\frac{E_2}{E_1} = \frac{K(s^2+8)}{(s+3)(s+8)}$$

find the network that yields this transfer function.

DIRECT SYNTHESIS METHODS

13-1. Introduction. As indicated in numerous places in the preceding chapters, a direct method of relating closed-loop performance characteristics to open-loop transfer functions is a desired goal in control system work. Direct synthesis is a procedure whereby a set of closed-loop specifications are interpreted in terms of a pole-zero pattern and a physically realizable open-loop transfer function is directly obtained from the closed-loop pole-zero pattern. This scheme contrasts sharply with the direct analysis and "synthesis by extended analysis" methods of Chapters 8, 9, and 11. These methods generally work in terms of the open-loop characteristics and attempt to achieve suitable closed-loop performance by indirect means such as the use of phase margin, M circles, and various other criteria. The root-locus schemes of Chapter 11 are about as direct an attack on system performance as any method considered in earlier chapters.

One useful direct synthesis scheme is the Guillemin method¹ for systems of the form of Fig. 13-1-1. This method follows the basic steps outlined below which in reality must apply to any similar direct synthesis method.

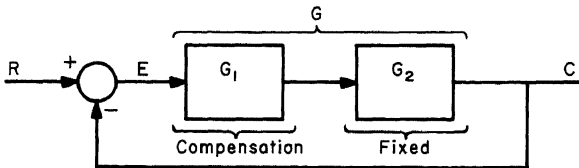


Fig. 13-1-1.

- (1) A closed-loop function $T(s) = C(s)/R(s)$ is formed from the over-all system specifications.
- (2) The open-loop transfer function $G(s) = C(s)/E(s)$ is determined from $T(s)$.
- (3) The necessary compensation function $G_1(s)$ is found from the transfer function for the fixed part of the system, $G_2(s)$, and the over-all open-loop transfer function, $G(s)$ as $G_1(s) = G(s)/G_2(s)$.

¹ John G. Truxal, *Automatic Feedback Control System Synthesis*, McGraw-Hill Book Co., Inc., New York, 1955, pp. 278-316.

- (4) The compensation function $G_1(s)$ is synthesized by network synthesis techniques (Chapter 12).

The above steps appear quite straight-forward; however between the lines are many hidden obstacles. For example, the compensation function $G_1(s)$ is generally required to be capable of realization by a network of resistances, capacitances, and an amplifier. This restriction also requires that the poles of $G_1(s)$ be on the negative real s -plane axis and that no poles exist at the origin or infinity as discussed in Art. 12-9. The complexity of the fixed part of the system, $G_2(s)$, must also be considered in choosing $T(s)$. In general the number of finite poles of $T(s)$ must exceed the number of finite zeros of $T(s)$ by at least the same pole-zero excess as contained in $G_2(s)$. Otherwise $G_1(s)$ will have a pole at infinity (Prob. 13-1). Of course $T(s)$ must also satisfy the system specifications. The details of the Guillemin method of synthesis account for the various restrictions but not without some rather involved steps. Considerable trial and error is required, if simplified compensation functions are to be realized.

A somewhat more direct synthesis approach for a Type 1 system of the form of Fig. 13-1-1 has been developed around a third-order system with a zero^{2,3}. This scheme is considered in detail in the following Articles.

13-2. Choice of Form for $T(s)$. As mentioned in Art. 13-1, $T(s)$ must meet system specifications and must normally have a certain pole-zero excess. However an excessive number of poles and zeros makes $T(s)$ completely unmanageable as far as direct control of such items as overshoot, bandwidth, rise time, etc. are concerned. Ideally a simple pair of poles as required for a basic second-order system (Chapter 6) are desirable. However the range of specifications which can be met by this system is quite limited as shown in Fig. 13-2-1. This figure is based on normalization of the equations

$$T(s) = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \quad (13-2-1)$$

$$G(s) = \frac{\omega_n^2}{s(s + 2\zeta\omega_n)} \quad (13-2-2)$$

as follows

$$T(p) = \frac{1}{p^2 + 2\zeta p + 1} \quad (13-2-3)$$

$$G(p) = \frac{1}{p(p + 2\zeta)} \quad (13-2-4)$$

² C. R. Hausenbauer and G. V. Lago, "Synthesis of Control Systems Based on Approximation to a Third-Order System," *Trans. AIEE* 77, Part II (November, 1958): 415-421.

³ C. R. Hausenbauer, *Synthesis of Feedback Systems*, Ph.D. thesis, University of Missouri, 1957.

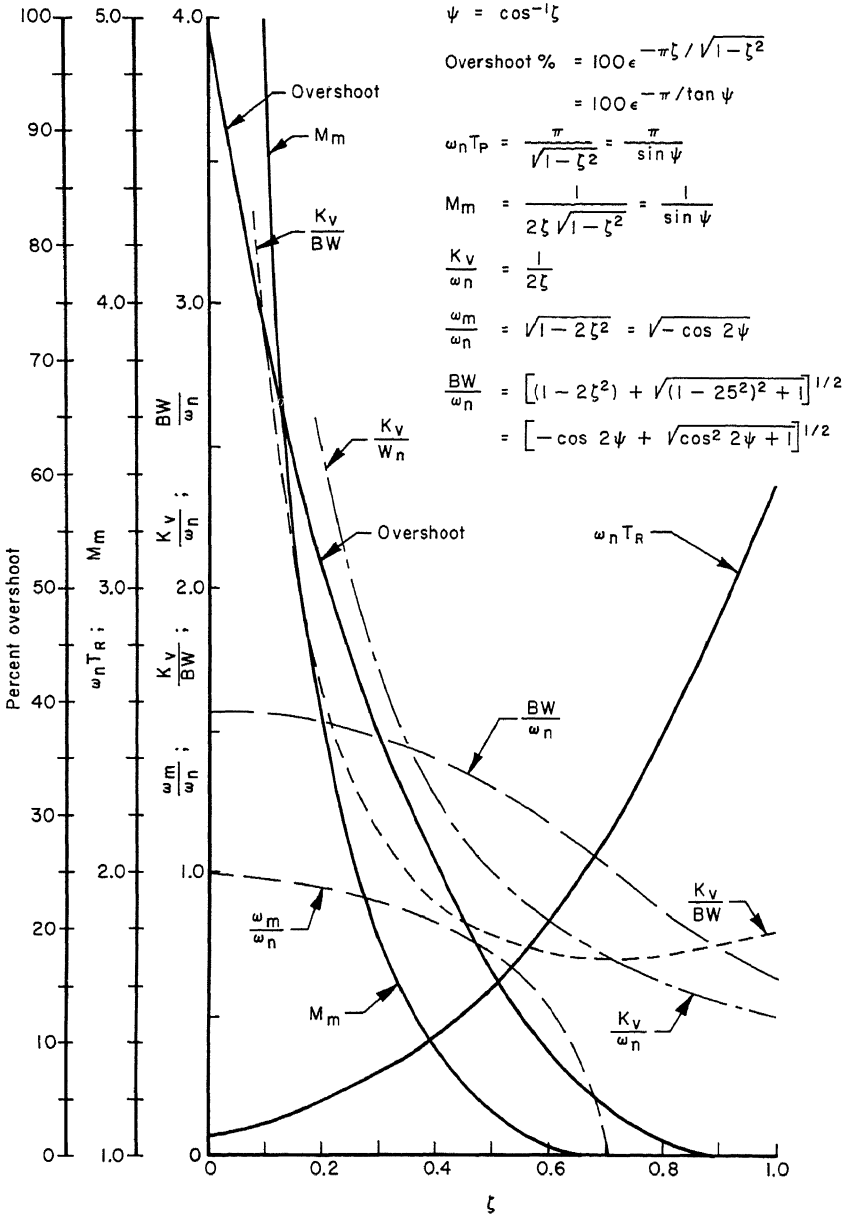


Fig. 13-2-1.

where $p = s/\omega_n$. In terms of bandwidth, BW , velocity constant, K_v , and time, t , in the original system, these quantities are now BW/ω_n , K_v/ω_n , and $t' = \omega_n t$. The s -plane is changed to a p -plane and the complex poles are on a unit radius circle. This normalization scheme is used at various places throughout the remainder of this chapter.

The addition of a simple zero to the basic second-order system extends the range of normalized velocity constant and bandwidth slightly but increases the overshoot and resonant peak for given ζ values as follows:

$$T(p) = \frac{p + k}{k(p^2 + 2\zeta p + 1)} \quad (13-2-5)$$

$$G(p) = \frac{p + k}{kp[p + (2\zeta - 1/k)]} \quad (13-2-6)$$

Fig. 10-7-15e presents typical transient responses. The effect of a zero on the normalized velocity constant is seen from Eq. 11-11-1

$$\frac{\omega_n}{K_v} = \left[\sum_i \frac{\omega_n}{\gamma_i} - \sum_i \frac{\omega_n}{\sigma_i} \right] \quad (13-2-7)$$

where γ_i is the negative of the unnormalized closed-loop poles and σ_i is the negative of the unnormalized closed-loop zeros. Any zero in the left half s -plane tends to lower ω_n/K_v or to raise K_v/ω_n . The increase in system bandwidth can be seen by considering phasors in the s -plane as s takes on values on the $j\omega$ -axis. As ω approaches k the phasor $j\omega + k$ counteracts to some extent the increase in denominator phasors, thereby causing the half-power frequency to be higher than in the basic second-order case. It might be noted that $k > 1/2\zeta$ for open-loop stable systems and that the pole-zero excess of this system is only one.

The addition of a simple pole to the basic second-order system provides very few benefits over a second-order system. The pole-zero excess is three. However for a fixed ζ for the complex pole pair the velocity constant is lowered as shown by Eq. 13-2-7, the bandwidth and overshoot are lowered, and the rise time is slower. A more detailed consideration of this system is reserved for problem work (Probs. 13-3 through 13-6.).

As a final form for $T(s)$ the third-order system with a zero is considered:

$$T(s) = \frac{\gamma_3 \omega_n^2}{\sigma_1} \frac{s + \sigma_1}{(s + \gamma_3)(s^2 + 2\zeta \omega_n s + \omega_n^2)} \quad (13-2-8)$$

$$T(p) = \frac{m}{k} \frac{p + k}{(p + m)(p^2 + 2\zeta p + 1)} \quad (13-2-9)$$

From the discussion of Chapter 10 and Fig. 10-7-15a to f, it is known that as long as $k < m$ this system has higher overshoot and faster rise time than a second-order system of the same ζ . From Eq. 13-2-7 it can be seen that a larger normalized velocity constant results, if $k < m$. For a considerable range of m and k the open-loop poles are real and on the negative real axis (Prob. 13-7). While the pole-zero excess is only two for the $T(s)$ of Eq. 13-2-8, the third-order system with a zero is the most versatile transfer function yet considered and it leads to manageable functional relations for the velocity constant, bandwidth, etc. Curves that summarize the per-

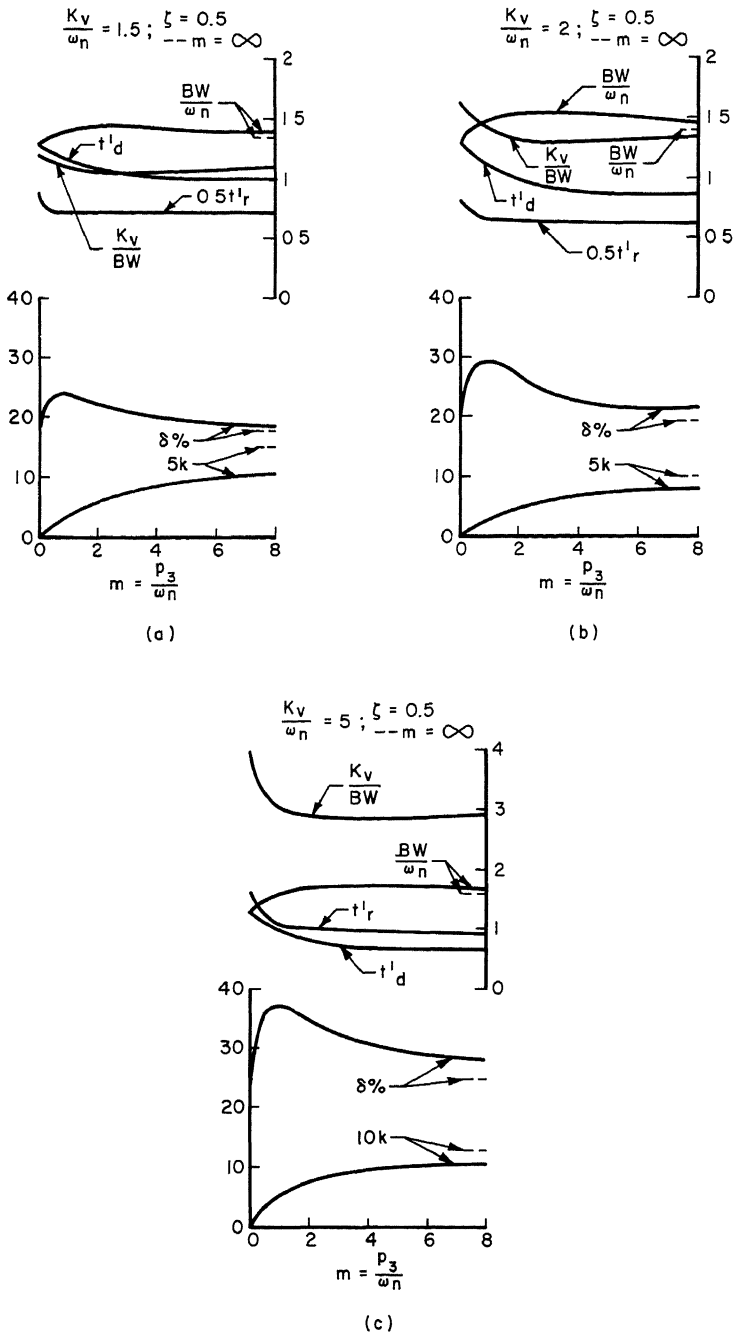


Fig. 13-2-2.

(From Trans. AIEE, Applications and Industry, with permission.)

performance capabilities of a system with the $T(p)$ of Eq. 13-2-9 are presented in Fig. 13-2-2 for constant values of K_v/ω_n with $\zeta = 0.5$ and in Fig. 13-2-3 for $\zeta = 0.707$. One new quantity appears in these figures. This is the v/BW ratio which is independent of ω_n . It is therefore, a convenient design parameter that can be used before an ω_n is chosen. Because of the wide range of specifications which can be met and because of the direct relations between specifications and pole-zero patterns, the third-order system with a zero is chosen as the basic form for $T(s)$ in the Hausenbauer-Lago synthesis procedure.

13-3. The Hausenbauer-Lago Synthesis Procedure. As in the Routh-Hurwitz synthesis procedure, the system of Fig. 13-1-1 is the basis of the Hausenbauer-Lago method. This method approaches the restrictions mentioned in Art. 13-1 in the following manner. The complexity of $G_2(s)$, the fixed part of the system, is ignored in forming $T(s)$. A normalized third-order system with a zero is chosen as the basic prototype (p) . This permits direct satisfaction of many sets of specifications (Figs. 13-2-2 and 13-2-3) but it also allows poles at infinity in the compensation function, $G_1(s)$. In order to cause $G_1(s)$ to be RC -realizable, additional open-loop poles are added so that the phase margin of $G(s)$ is not adversely affected. The condition equation

$$\frac{m}{1 + 2\zeta m} \leq k \leq \frac{m}{1 - (\zeta - m/2)^2} \quad (13-3-1)$$

and the $T(p)$ of Eq. 13-2-8 is used to assure that the poles of $G(p)$ are on the negative real p - and s -plane axis. No other difficulties in obtaining an RC -realizable compensation result for a Type 1 system.

A consideration of Step (3) of the basic synthesis procedure outlined in Art. 13-1 indicates a way of simplifying the compensation functions, $G_1(s)$. From the relation

$$G_1(s) = G(s)/G_2(s) \quad (13-3-2)$$

it can be seen that any poles and zeros common to $G(s)$ and $G_2(s)$ can be canceled. Therefore, if $G(s)$ can be deliberately formed to contain some singularities of $G_2(s)$, the compensation function $G_1(s)$ can be simplified. For the third-order system with a zero the open-loop transfer function is

$$G(p) = \frac{m}{k} \frac{p + k}{p(p + p_4)(p + p_5)} \quad (13-3-3)$$

where m and k are defined by the closed-loop relation of Eq. 13-2-8 repeated as

$$T(p) = \frac{m}{k} \frac{p + k}{(p + m)(p^2 + 2\zeta p + 1)} \quad (13-3-4)$$

The positions of p_4 and p_5 in Eq. 13-3-3 can be varied over a reasonable range while still allowing $T(p)$ to meet various specifications. Thus there

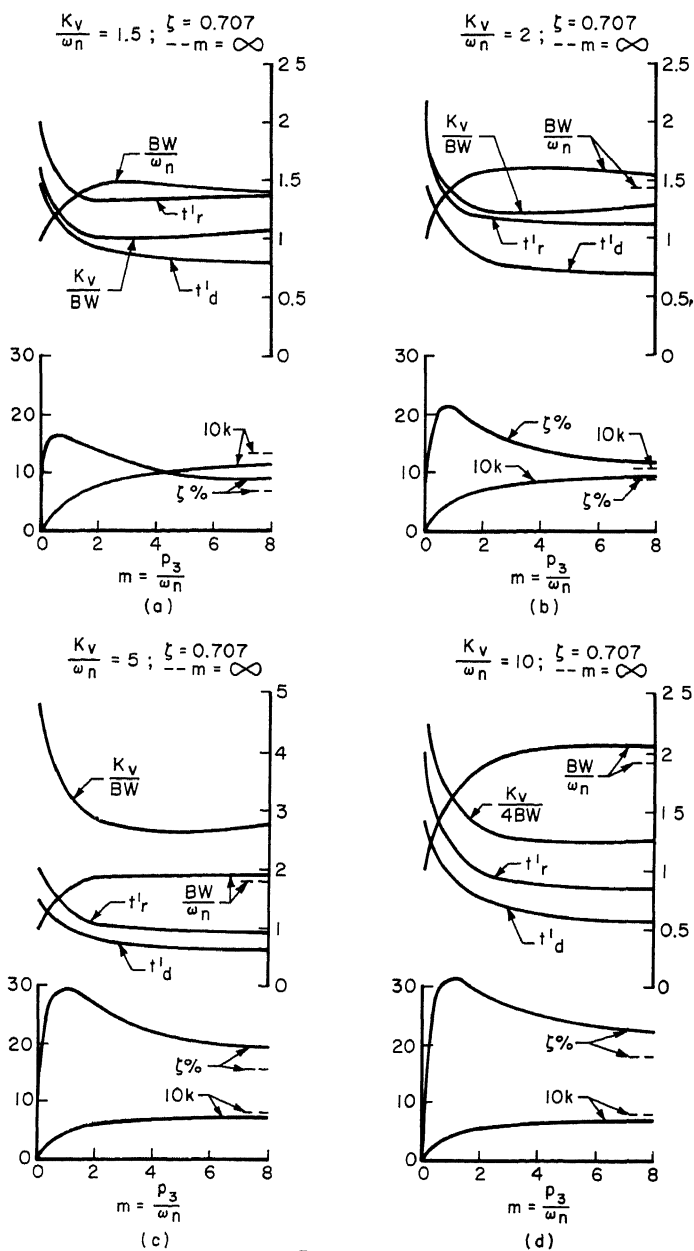


Fig. 13-2-3.

(From Trans. AIEE, Applications and Industry, with permission.)

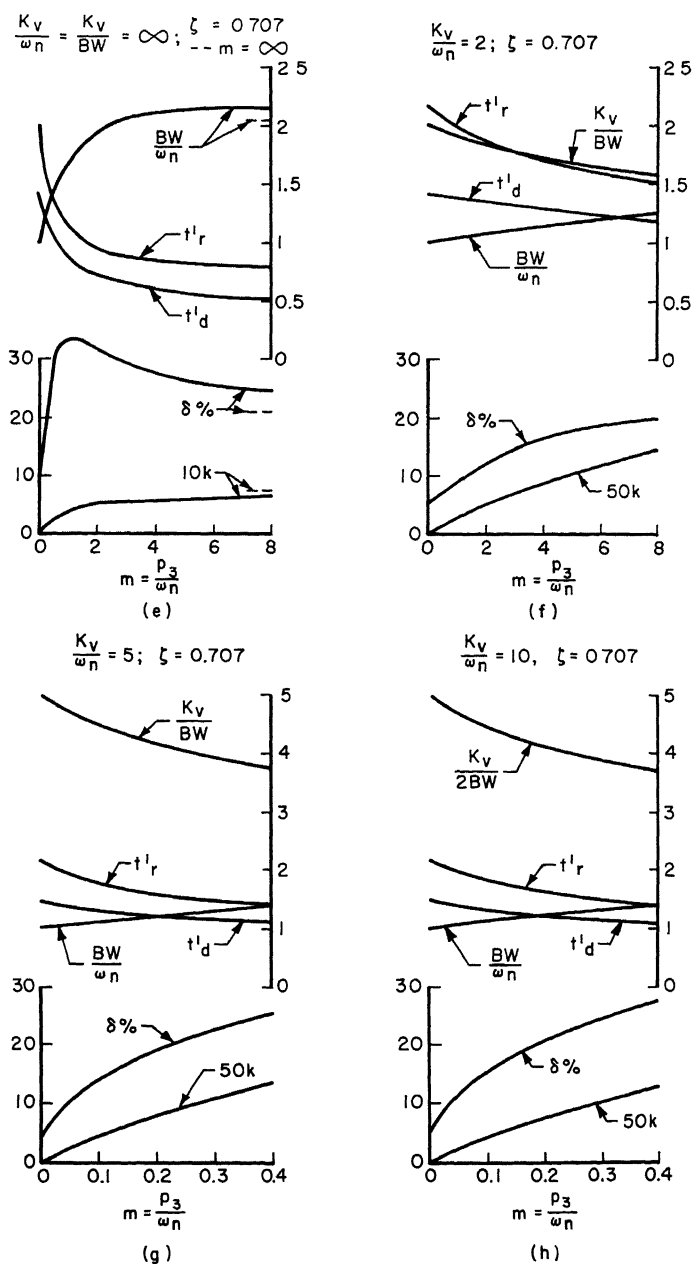


Fig. 13-2-3 (cont.).

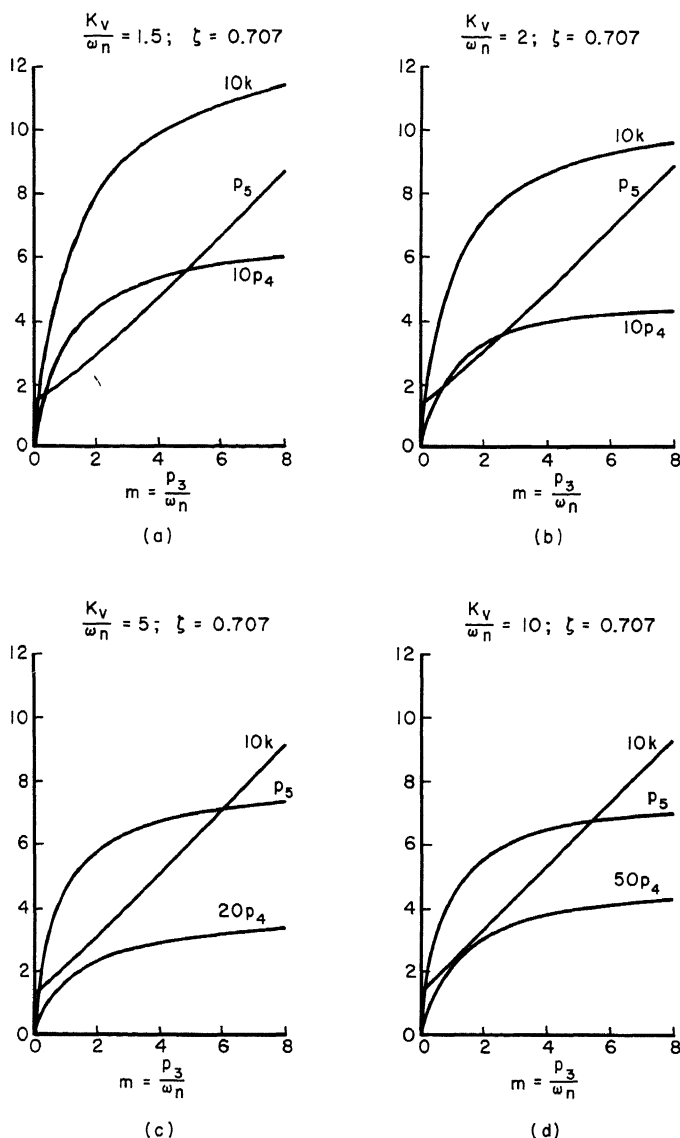


Fig. 13-3-1.

(From Trans. AIEE, Applications and Industry, with permission.)

is considerable freedom in moving p_4 or p_5 to cancel fixed open-loop poles and give simplified compensation functions. Fig. 13-3-1 presents curves that allow direct consideration of the positions of p_4 and p_5 as functions of the closed-loop pole m for $\zeta = 0.707$.

The actual steps of the Hausenbauer-Lago synthesis procedure can now be outlined as follows.

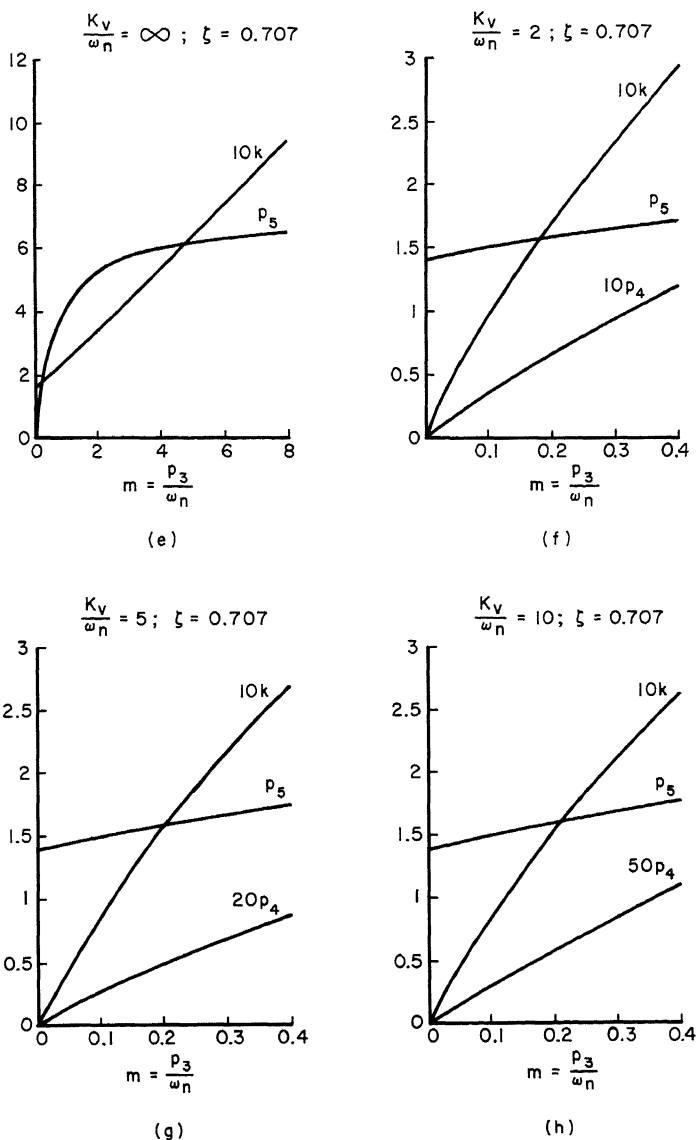


Fig. 13-3-1 (cont.).

- (1) The K_v/BW ratio is formed and used to determine a suitable set of design curves from Figs. 13-2-2 and 13-2-3 and a suitable value of ω_n based on the required BW .
- (2) $G_2(s)$ is normalized to $G_2(p)$ using ω_n from Step (1).
- (3) A $T(p)$ and a $G(p)$ are formed using the design curves of Fig. 13-3-1 to try to obtain pole cancellation between $G(p)$ and $G_2(p)$ while simultaneously using the specification curves of Figs. 13-2-2 and 13-2-3.

- (4) The idealized compensation function $G_1(p)$ is determined from $G(p)/G_2(p)$.
- (5) Since $G_1(p)$ of Step (4) may not be RC -realizable because of poles at infinity, remote poles are added using a Bode plot as a guide so that the phase margin is not seriously affected.
- (6) The final modified $G_1(p)$ is synthesized as an RC network using the procedures of Chapter 12 and unnormalized with $p = s/\omega_n$ to obtain the final compensation function.
- (7) The actual final $T(p)$ is checked if any doubt exists as to its meeting the original specifications.

The details of carrying out the above procedure are illustrated in the examples of the following article.

13-4. Examples of the Hausenbauer-Lago Synthesis Procedure.

The following examples not only illustrate the synthesis procedure under discussion, but also provide an application of the RC network synthesis procedures of Chapter 12.

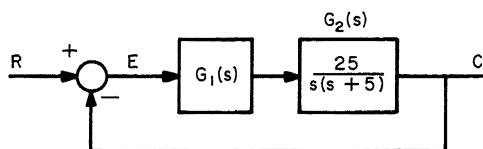


Fig. 13-4-1.

Example 13-4-1. A Type 1 unity feedback system with the fixed part of the system as shown in Fig. 13-4-1 is to be redesigned to meet the following specifications:

$$K_v \geq 100 \quad ; \quad BW \leq 40 \text{ radians/sec} \quad ; \quad \text{Overshoot} \leq 20\%$$

As a preliminary step the $G_1(s)$ is set equal to unity to see what conditions exist in the basic system by itself. The following conditions are determined mainly from Fig. 13-2-1.

$$\begin{array}{ll} \omega_n = 5 \text{ radians/sec} & BW = 6.4 \text{ radians/sec} \\ \zeta = 0.5 & K_v/BW = 0.78 \\ K_v = 5 \text{ sec}^{-1} & \text{Overshoot} = 16.5\% \end{array}$$

Thus a factor of 20 times improvement in K_v has to be obtained along with a wider bandwidth and little change in the overshoot. For the desired specifications, $K_v/BW \geq 2.5$. A preliminary study of the design curves of Fig. 13-2-2 and 13-2-3 indicates that $K_v/\omega_n = 5$ and $\zeta = 0.707$ meet the requirements if $m > 5.6$. For most of the remaining figures either K_v/BW is too low or the overshoot is too high. For a choice of $K_v/\omega_n = 5$, $\zeta = 0.707$, $\omega_n = 20$ provides that $K_v = 100$ and $BW = 40$ radians/sec. For this ω_n the fixed part of the system is normalized as

$$G_2(p) = \frac{1/16}{p(p + 0.25)} \quad (13-4-1)$$

The curves of Fig. 13-3-1 indicate no possibility of canceling the open-loop fixed pole at $p = -0.25$ and so this allows m to be chosen subject only to specifications.

The following values are determined from Fig. 13-2-3 for $m = 8.0$:

$$\begin{array}{ll} 10k = 7.46 & p_s = 9.24 \\ K_v/BW = 2.7 & 20p_d = 3.45 \\ BW/\omega_n = 1.9 & \text{Overshoot} = 18.5\% \\ t'_r = 0.88 & K_v/\omega_n = 5 \\ t'_d = 0.57 & \omega_n = 20 \\ \zeta = 0.707 & K_v = 100 \end{array}$$

The normalized closed-loop transfer function is

$$T(p) = \frac{m}{k} \frac{p + k}{(p + m)(p^2 + 2\zeta p + 1)} \quad (13-4-2)$$

$$= 10.71 \frac{p + 0.746}{(p + 8)(p^2 + \sqrt{2}p + 1)} \quad (13-4-3)$$

The normalized open-loop transfer function is

$$G(p) = \frac{(10.71)(p + 0.746)}{p(p + 0.173)(p + 9.24)} \quad (13-4-4)$$

Eq. 13-4-1 allows the compensation function to be determined as

$$G_1(p) = \frac{G(p)}{G_2(p)} = \frac{(171.2)(p + 0.25)(p + 0.746)}{(p + 0.173)(p + 9.24)} \quad (13-4-5)$$

As indicated above, no pole cancellations occur, and the compensation function is rather involved but certainly realizable by an RC network (assuming part of the system is electrical). Either $G_1(p)$ or $G_1(s)$ can be synthesized with no difference in results. If $G_1(p)$ is synthesized, the unnormalization process requires that all capacitors be divided by ω_n since the operational admittance of a capacitor is

$$Cp = C \frac{s}{\omega_n} = \frac{C}{\omega_n} s \quad (13-4-6)$$

The transfer function $G_1(p)$ is synthesized as a voltage ratio using $G_1(p) = y_{12}/y_{22}$ from Art. 12-9. Thus

$$y_{12}(p) = \frac{K(p + 0.25)(p + 0.746)}{q(p)} \quad (13-4-7)$$

$$y_{22}(p) = \frac{(p + 0.173)(p + 9.24)}{q(p)} \quad (13-4-8)$$

Since $y_{22}(p)$ must have a zero inside its pole closest to the origin, and since $y_{12}(p)$ is simplified, one pole is chosen at $p = -0.25$. Another pole is arbitrarily chosen at -15 so that loading at both ends of $G_1(p)$ can be realized. This results in

$$y_{12}(p) = \frac{K(p + 0.746)}{p + 15} \quad (13-4-9)$$

$$y_{22}(p) = \frac{(p + 0.173)(p + 9.24)}{(p + 0.25)(p + 15)} \quad (13-4-10)$$

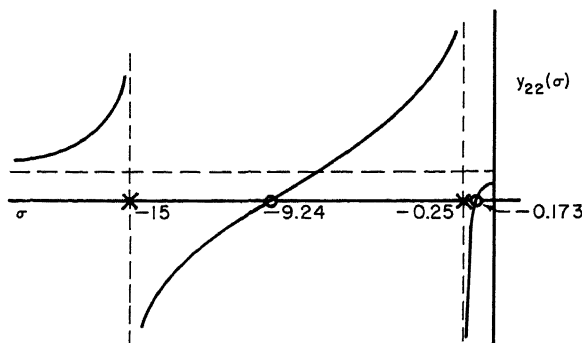


Fig. 13-4-2.

Fig. 13-4-2 shows a sketch of y_{22} for values of p along the negative real axis. The pole at -0.25 must be removed as a shunt branch if extra zeros of transmission in y_{12} are to be avoided. Also a shunt resistance must be removed for loading purposes. The pole at -0.25 in $y_{22}(p)/p$ is removed, since its removal does not affect the ability to remove a shunt resistance branch later.

$$\frac{y'_{22}(p)}{p} = \frac{y_{22}(p)}{p} - \frac{k_{-0.25}}{p + 0.25} \quad (13-4-11)$$

$$= \frac{1}{p + 0.25} \left[\frac{(p + 0.173)(p + 9.24)}{p(p + 15)} - 0.1877 \right]$$

$$y'_{22}(p) = \frac{(0.8123)(p + 7.86)}{p + 15} \quad (13-4-12)$$

An admittance branch

$$y_a = \frac{0.1877p}{p + 0.25} \quad (13-4-13)$$

is removed from y_{22} by the above operation. Fig. 13-4-3a shows the partial network thus far.

The zero in y_{22} at -7.86 can be moved to -0.746 by removal of a shunt conductance for loading purposes and also to provide the zero in y_{12} .

$$y'_{22}(-0.746) = \frac{(0.8123)(7.86 - 0.746)}{15 - 0.746} = 0.4054 \quad (13-4-14)$$

$$y''_{22}(p) = y'_{22}(p) - 0.4054 = \frac{(0.4069)(p + 0.746)}{p + 15} \quad (13-4-15)$$

Fig. 13-4-3b shows the partial network after removal of the two shunt branches. To realize the zero in y_{12} at -0.746 , $y''_{22}(p)$ is inverted and a series impedance branch is removed as the pole in $1/y''_{22}(p)$ at -0.746 .

$$\frac{1}{y''_{22}(p)} = \frac{(p + 15)}{(0.4069)(p + 0.746)} = \frac{1}{0.4069} + \frac{k_{-0.746}}{p + 0.746} \quad (13-4-16)$$

$$k_{-0.746} = \frac{14.25}{0.4069} \quad (13-4-17)$$

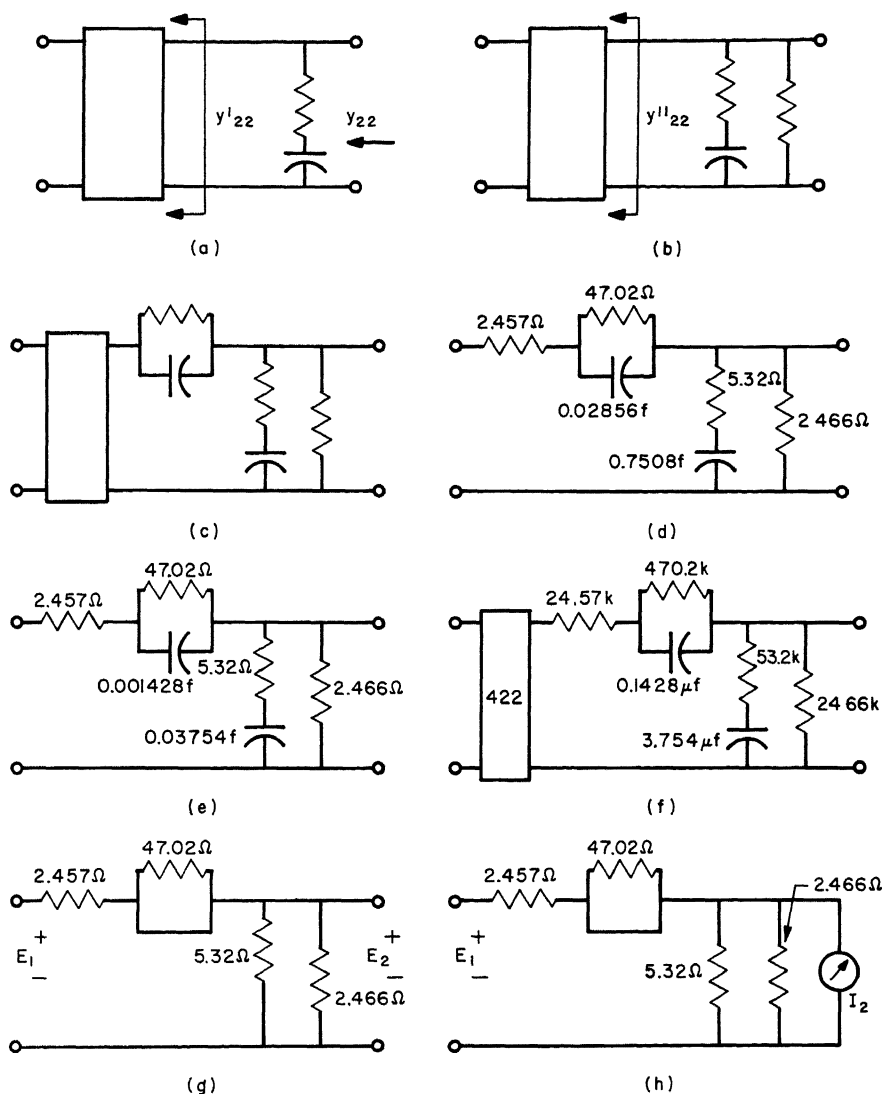


Fig. 13-4-3.

The pole removal provides parallel RC elements in a series arm and the remaining resistance placed in series for loading. Figs. 13-4-3c and d show the final normalized network. The unnormalized network is obtained by dividing all capacitors by $\omega_n = 20$ to give the network of Fig. 13-4-3e. Since y_{12} is only realized within a gain constant K , this constant must be determined. From Eq. 13-4-9 as p approaches infinity, y_{12} approaches K . All capacitors in the network approach short circuits and by definition y_{12} is found from Fig. 13-4-3h as

$$y_{12} = \left. \frac{I_2}{E_1} \right|_{E_2=0} \rightarrow \frac{1}{2.457} = 0.407 \quad (13-4-18)$$

Thus $K = 0.407$. The gain constant can also be found at d-c from the E_2/E_1 relation and Fig. 13-4-4g. With the above K an ideal amplifier must be used to obtain the gain constant 171.2 in Eq. 13-4-5. The gain required is

$$A = \frac{171.2}{0.407} = 420.6 \quad (13-4-19)$$

as shown in Fig. 13-4-3f which also shows the impedance level of the network scaled up by a factor of 10^4 . This last scaling operation provides reasonable component values and does not alter time constants or gains, since all resistors are multiplied by 10^4 and all capacitors are divided by this factor.

Several facts should be pointed out before leaving this example. The basic network provides modified lag-lead compensation. Since the fixed part of the system had a pole-zero excess of two as did $T(p)$, no pole occurred at infinity in $G_1(p)$. Thus no remote poles had to be added to make $G_1(p)$ RC-realizable.

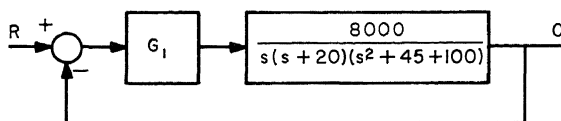


Fig. 13-4-4.

Example 13-4-2. As a second example of the Hausenbauer-Lago synthesis procedure, the system of Fig. 13-4-4 is considered. The requirements for this system are $K_v \geq 50 \text{ sec}^{-1}$, $BW \leq 30 \text{ radians/sec}$, overshoot $\leq 30\%$. For the basic system $K_v = 4$ and the rough root-locus plot of Fig. 13-4-5 shows the system

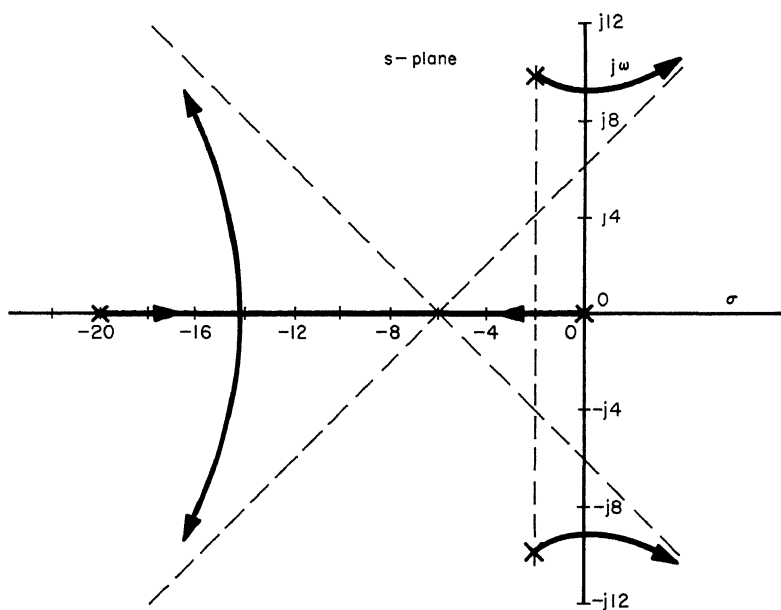


Fig. 13-4-5.

to be very near instability. The specifications require $K_v/BW \geq 1.67$. Of the several design curves in Figs. 13-2-2 and 13-2-3 which are suitable, the one for $K_v/\omega_n = 5$, $\zeta = 0.707$ is chosen again. For this case $BW/\omega_n < 2$, so that an $\omega_n = 10$ radians/sec is suitable. Normalization of $G_2(s)$ results in

$$G_2(p) = \frac{0.8}{p(p+2)(p^2+0.4p+1)} \quad (13-4-20)$$

a check of the curves of Fig. 13-3-1 indicates a possibility of including the pole at -2 in $G(p)$ if m is small. By repeated trials an $m = 0.652$ is found suitable for which the remaining system quantities are

$k = 0.364$	$t'_d = 1.05$
$p_5 = 2.0$	Overshoot = 27.8%
$p_4 = 0.0652$	$K_v/\omega_n = 5$
$K_v/BW = 3.45$	$\omega_n = 10$ radians/sec
$BW/\omega_n = 1.45$	$BW = 14.5$ radians/sec
$t'_r = 1.3$	

Thus the required transfer functions are

$$T(p) = \frac{(1.79)(p+0.364)}{(p+0.652)(p^2+\sqrt{2}p+1)} \quad (13-4-21)$$

$$G(p) = \frac{(1.79)(p+0.364)}{p(p+2)(p+0.0652)} \quad (13-4-22)$$

$$G_1(p) = \frac{(2.24)(p+0.364)(p^2+0.4p+1)}{p+0.0652} \quad (13-4-23)$$

In order to obtain an RC -realizable compensation two remote poles must be added along the negative real axis. A Bode plot of $G(p)$ indicates crossover in the vicinity of 0.85 dimensionless frequency units with a phase margin of 48° . Since the overshoot is already nearly 30 per cent, the poles should contribute very little phase lag at this frequency. To show the effect on overshoot, poles that are marginal are added at $p = -15$ and -25 with a decrease of 5.2° in the phase margin. A detailed check shows that the final 42.8° phase margin results in about 33 per cent overshoot. Nevertheless the above poles are retained. The modified compensation function

$$G'_1(p) = \frac{(15)(25)(2.24)(p+0.364)(p^2+0.4p+1)}{(p+0.0652)(p+15)(p+25)} \quad (13-4-24)$$

is again realized as a voltage ratio

$$\frac{E_2}{E_1} = G'_1(p) = \frac{y_{12}}{y_{22}} \quad (13-4-25)$$

The complex zeros suggest the use of the parallel ladder development of Art. 12-10. For the over-all network

$$y_{12} = \frac{K(p+0.364)(p^2+0.4p+1)}{q(p)} \quad (13-4-26)$$

$$y_{22} = \frac{(p+0.0652)(p+15)(p+25)}{q(p)} \quad (13-4-27)$$

from which a suitable $q(p)$ is

$$q(p) = (p + 0.364)(p + 20) \quad (13-4-28)$$

A split up of y_{12} into α and β parts which are realizable by ladders is

$$y_{12} = \underbrace{\frac{K_a p^2}{p + 20}}_{\alpha} + \underbrace{\frac{K_\beta (0.4)(p + 2.5)}{p + 20}}_{\beta} \quad (13-4-29)$$

The α network can be realized by a use of Cauer's second form after removal of a pole term to prevent a triple zero in y_{12} . The pole at -0.364 is removed from y_{22}/p as follows.

$$\frac{y'_{22\alpha}}{p} = \frac{y_{22\alpha}}{p} - \frac{15.07}{p + 0.364} = \frac{(p + 21.5)(p + 3.14)}{p(p + 20)} \quad (13-4-30)$$

$y'_{22\alpha}$ is now developed as

$$y'_{22\alpha} = \frac{67.25}{20} + \frac{1}{\frac{20}{21.27p} + \frac{1}{\frac{21.27}{1.27} + \frac{1}{1.27/p}}} \quad (13-4-31)$$

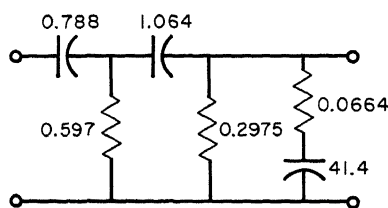


Fig. 13-4-6.

to give the network of Fig. 13-4-6. The β network requires only a single zero at -2.5 . The expansion of $y_{22\beta}$ is started in a manner similar to the $y_{22\alpha}$ expansion by removing the pole at -0.364 from $y_{22\beta}/p$. An excess of zeros still exists thereby indicating the removal of a shunt capacitance as follows.

$$y''_{22\beta} = y'_{22\beta} - p = \frac{(p + 21.5)(p + 3.14)}{p + 20} - p \quad (13-4-32)$$

$$= \frac{4.63p + 67.25}{p + 20} \quad (13-4-33)$$

The zero in $y''_{22\beta}$ can now be moved to -2.5 by conductance removal.

$$y'''_{22\beta} = y''_{22\beta} - 3.181 = \frac{1.45(p + 2.5)}{p + 20} \quad (13-4-34)$$

The zero in $y_{12\beta}$ at -2.5 is now obtained by inverting $y'''_{22\beta}$ and removing the pole in $1/y'''_{22\beta}$ at -2.5 .

$$\frac{1}{y'''_{22\beta}} = \frac{1}{1.45} \left(1 + \frac{17.5}{p + 2.5} \right) \quad (13-4-35)$$

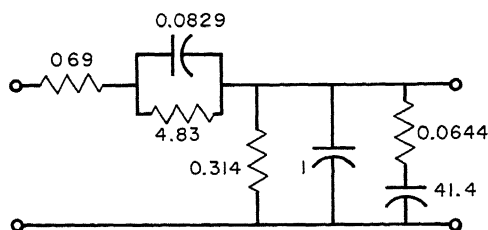
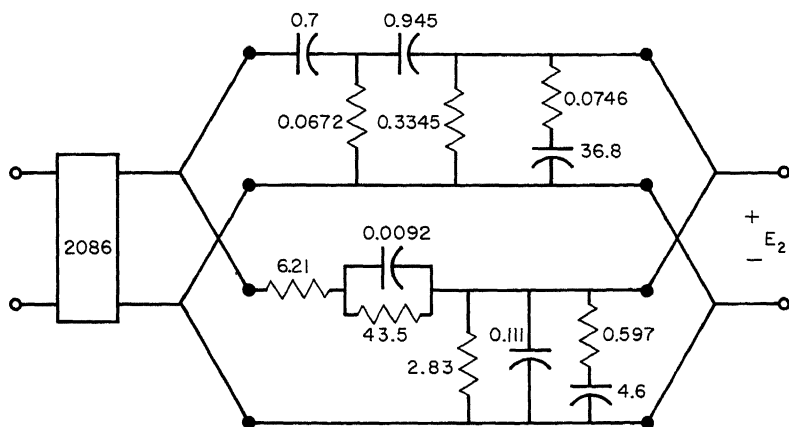
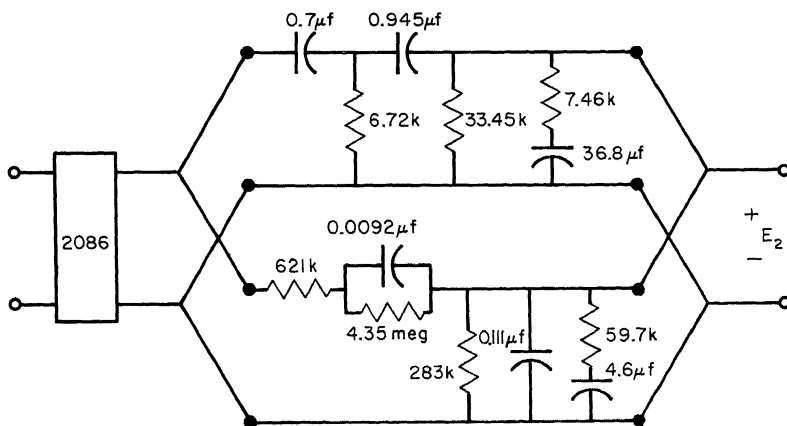


Fig. 13-4-7.

The remaining resistance, $1/1.45$, is also placed in series as shown in the network of Fig. 13-4-7. The a and β gain constants K_a and K_β must now be determined so that final scaling of y_{22} 's can be accomplished. From Fig. 13-4-6 at very high frequencies and Eq. 13-4-29,



(a)



(b)

Fig. 13-4-8.

$$y_{12a} \rightarrow K_a p = Cp = \frac{(0.788)(1.064)}{0.788 + 1.064} p \quad (13-4-36)$$

$$K_a = 0.453 \quad (13-4-37)$$

Similarly for K_β

$$y_{12\beta} \rightarrow 0.4K_\beta = 1/0.69 = 1.45 \quad (13-4-38)$$

$$K_\beta = 3.625 \quad (13-4-39)$$

Parallel connection of the α and β networks requires scaling of each network so that y_{12} gain constants are equal and so that the total y_{22} is that given by Eq. 13-4-27. Thus

$$\alpha K_a = \beta K_\beta \quad ; \quad \alpha + \beta = 1 \quad (13-4-40)$$

$$0.453\alpha = 3.63\beta \quad ; \quad \alpha = 0.889, \beta = 0.111 \quad (13-4-41)$$

Scaling of α network admittances by 0.889 and β network admittances by 0.111 gives the final normalized network of Fig. 13-4-8a. The amplification required is determined from the gain in Eq. 13-4-24 and the over-all y_{12} constant αK_a or βK_β

$$A = \frac{(15)(25)(2.24)}{(0.111)(3.63)} = 2086 \quad (13-4-42)$$

Division of all capacitors by $\omega_n = 10$ converts the network to a real time base and indicates a final impedance level change of 10^5 for reasonable component values. The final scaled network is shown in Fig. 13-4-8b.

The above examples illustrate the Hausenbauer-Lago synthesis procedure for Type 1 unity feedback control systems. For such systems it provides a direct design procedure with immediate correlation between specifications and transfer functions.

PROBLEMS

13-1. Show that $T(s) = C(s)/R(s)$ must have the same or greater pole-zero excess as contained in $G_2(s)$, if $G_1(s)$ is not to contain a pole at infinity. Use two methods: (a) consider limits as s approaches infinity of the expression

$$T(s) = \frac{G_1(s)G_2(s)}{1 + G_1(s)G_2(s)}$$

(b) consider $G_1(s)$ and $G_2(s)$ as ratios of polynomials and construct $T(s)$.

13-2. Develop an expression for the unit-step response of the system of Eq. 13-2-5 in terms of $t' = \omega_n t$ and sketch the response for $k = 2$, $\zeta = 0.4$. Determine the normalized velocity constant K_v/ω_n for this case. Compare the results with a basic second-order system with $\zeta = 0.4$.

13-3. Using a closed-loop transfer function of the form

$$\frac{C(s)}{R(s)} = \frac{\gamma_3 \omega_n^2}{(s + \gamma_3)(s^2 + 2\zeta \omega_n s + \omega_n^2)} = \frac{G(s)}{1 + G(s)}$$

normalize this so that the complex poles lie on a unit radius circle. Determine the restrictions on γ_3/ω_n for the poles of $G(p)$ to be on the negative real axis.

13-4. By use of the conventional velocity constant definition

$$K_v = \lim_{s \rightarrow 0} sG(s)$$

show that the same answer is obtained for the system of Prob. 13-3 as by using Eq. 13-2-7.

13-5. For a $\zeta = 0.5$ and $\gamma_3/\omega_n = 2$ find the step response of the system of Prob. 13-3. Compare this with a second-order system with $\zeta = 0.5$.

13-6. Construct a Bode plot of the closed-loop transfer function of Prob. 13-3 for $\zeta = 0.5$ and $\gamma_3/\omega_n = 2$. Determine BW from this plot and compare with the BW for a second-order system with $\zeta = 0.5$.

13-7. For

$$T(p) = \frac{G(p)}{1 + G(p)} = \frac{m}{k} \frac{p + k}{(p + m)(p^2 + 2\zeta p + 1)}$$

determine $G(p)$. What condition equation is imposed on m/k for the poles of $G(p)$ to fall on the negative real axis?

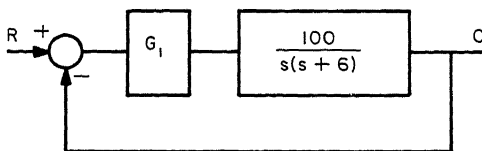


Fig. 13-P-8.

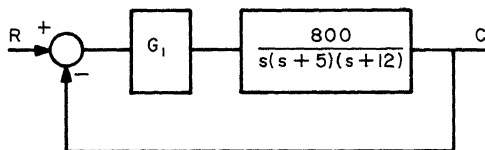


Fig. 13-P-9.

13-8. Using the design curves of Figs. 13-2-2, 13-2-3, and 13-3-1, design a control system based on Fig. 13-P-8 that will have

$$K_v \geq 40 \text{ sec}^{-1}$$

$$BW \leq 30 \text{ radians/sec}$$

$$t_r \leq 0.1 \text{ sec}$$

$$\text{Overshoot} \leq 25\%$$

Synthesize the required compensation function as a voltage ratio using an R - C network. Include input and output loading and scale all elements to reasonable sizes ($C < 100\mu\text{f}$, $R < 10 \text{ megohm}$).

13-9. Repeat Prob. 13-8 except use

$$K_v \geq 30 \text{ sec}^{-1}$$

$$t_r \leq 0.05 \text{ sec}$$

$$\text{Overshoot} \leq 10\%$$

and the basic system of Fig. 13-P-9.

13-10. Repeat Prob. 13-8 except use

$$K_v \geq 20 \text{ sec}^{-1}$$

$$\text{Overshoot} \leq 25\%$$

$$BW \leq 20 \text{ radians/sec}$$

$$t_r \leq 0.1 \text{ sec}$$

$$G_2(s) = \frac{5}{s(s^2 + 1.6s + 4)}$$

Neglect input loading in synthesizing the compensation function.

ANALOGS AND ANALOG COMPUTERS

14-1. Introduction. The response of many physically different systems is characterized by differential equations which are similar in form. On this basis mechanical, electrical, hydraulic, pneumatic, and many other types of systems can be considered as having certain analogies. Thus in one scheme mechanical force is analogous to electrical voltage as is shown in Art. 14-2. By using such analogs, knowledge in one field can be brought to bear on problems in another field. Also, easily constructed electrical circuits can be used to study the behavior of not-so-easily constructed mechanical or hydraulic devices.

A slightly different use of analogs occurs in electronic analog computers. These devices actually solve system differential equations by using various voltages as analogs for the dependent variables in the system equations. Such a scheme has great possibilities, since it can be used to perform nonlinear operations as well as linear ones. The restriction to linear approximations which is used in previous chapters is thus removed, and more accurate studies can be made. Increased complexity can also be handled with relative ease. Thus analogs and analog computers are welcome additions to the control engineer's family of tools for analysis and synthesis. The following articles describe some of the basic system analogs and electronic analog computer operations.

14-2. Mechanical-Electrical Analogs. The Force (Torque)-Voltage Analog. Two forms for analogs between mechanical systems and electrical circuits are considered in this book. The first of these systems is termed the force (torque)-voltage analogy which is investigated

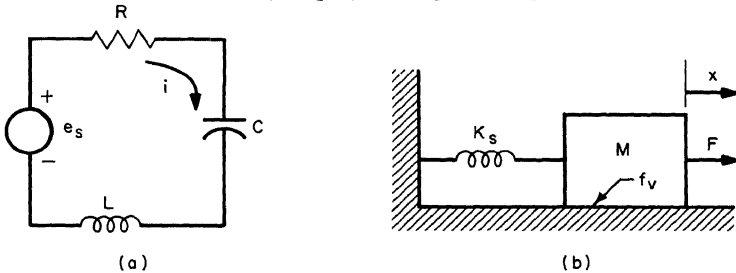


Fig. 14-2-1.

in this article. This analogy is established by using Fig. 14-2-1a and b instead of the individual parameter equations of Chapters 3 and 4. For the simple series circuit of Fig. 14-2-1a Kirchhoff's voltage law leads to the equation

$$L \frac{di}{dt} + iR + \frac{1}{C} \int_0^t idt = e_s \quad (14-2-1)$$

In terms of electric charge, $q = \int_0^t idt$, Eq. 14-2-1 can be written as

$$L \frac{d^2q}{dt^2} + R \frac{dq}{dt} + \frac{q}{C} = e_s \quad (14-2-2)$$

For the mechanical system of Fig. 14-2-1b the equation defining motion is

$$M \frac{d^2x}{dt^2} + f_v \frac{dx}{dt} + K_s x = F \quad (14-2-3)$$

The "equality of form" between Eqs. 14-2-2 and 14-2-3 is readily apparent, thus facilitating identification of one set of analogous quantities

TABLE 14-2-1

Electrical		Mechanical			
		Translational		Rotational	
Voltage	e	Force	F	Torque	T
Charge	q	Linear displacement	x	Angular displacement	θ
Current	i	Velocity	dx/dt	Angular velocity	ω
Inductance	L	Mass	M	Moment of inertia	J
Resistance	R	Viscous friction coefficient	f_v	Viscous friction coefficient	D
Elastance (reciprocal capacitance)	$1/C$	Spring constant	K_s	Spring constant	K_s

for mechanical and electrical systems. Table 14-2-1 presents a summary of these analogous quantities for the force (torque)-voltage system. Use of Table 14-2-1 makes it easy to substitute parameters and variables in system equations; however the problem of numerical relations is more involved. This problem of numerical values and consistent sets of units is considered in the following examples.

Example 14-2-1. An electrical circuit is to be constructed to study a mechanical system of the form of Fig. 14-2-1b. The mechanical parameters are $M = 0.5$ slug, $f_v = 0.5$ lb/(ft/sec), $K_s = 0.5$ lb/ft, and $F = 0.5$ lb. One possible analogy would have force in pounds represented by voltage in volts and displacement in feet analogous to charge in coulombs. If there is no change in the time variable between systems, then $L = 0.5$ henry, $R = 0.4$ ohm, $C = 2$ farads. The above scheme has several disadvantages resulting from component size and time duration of transients. A quick analysis indicates a natural angular velocity of 1 radian/sec so that the period of an undamped oscillation is 6.28 sec. If this is speeded up, an oscilloscope can be used to study the system response. Another point of concern is the amount of charge representing a foot of displacement. The basic problem is really one of *amplitude* and *time scaling*. The quantities in the analogy can be rescaled as follows. Each term in the mechanical equation is a force. Thus

$$\underbrace{M \frac{d^2 x}{dt^2}}_{\substack{\text{Accelerating} \\ \text{force}}} + \underbrace{f_v \frac{dx}{dt}}_{\substack{\text{Viscous} \\ \text{drag} \\ \text{force}}} + \underbrace{K_s x}_{\substack{\text{Spring} \\ \text{force}}} = \underbrace{F}_{\substack{\text{Applied} \\ \text{force}}} \quad (14-2-4)$$

Each such term is analogous to a voltage term in the electrical equation

$$L \frac{d^2 q}{d\tau^2} + R \frac{dq}{d\tau} + \frac{q}{C} = e_s(\tau) \quad (14-2-5)$$

where τ is time in the electrical circuit. Three quantities can be scaled—force, displacement (velocity or acceleration), and time. The general scaling problem can be illustrated as follows, using the viscous friction force term and the voltage drop across the resistance. The analogy is indicated by

$$\begin{array}{ccc} & \downarrow & \downarrow \\ f_v \frac{dx}{dt} \text{ lbs} & \longleftrightarrow & R \frac{dq}{d\tau} \text{ volts} \\ & \uparrow & \uparrow \end{array} \quad (14-2-6)$$

A scale factor a_f lbs/volt between force and voltage allows the following equation to be written.

$$\left(f_v \frac{dx}{dt} \right) \left(\frac{1}{a_f} \right) = R \frac{dq}{d\tau} \quad (14-2-7)$$

Next a scale factor a_t electrical seconds/mechanical second requires $\tau = a_t t$. The substitution of this relation gives

$$\left(\frac{f_v a_t}{a_f} \right) \frac{dx}{d\tau} = R \frac{dq}{d\tau} \quad (14-2-8)$$

Finally a scale factor a_x feet/coulomb establishes the relation $x = a_x q$ which results in

$$\left(\frac{a_x a_t f_v}{a_f} \right) \frac{dq}{d\tau} = R \frac{dq}{d\tau} \quad (14-2-9)$$

$$R = \left(\frac{a_x a_t f_v}{a_f} \right) \frac{\text{volt-electrical seconds}}{\text{coulomb}} \text{ or ohms} \quad (14-2-10)$$

By a similar process

$$L = \frac{a_x a_t^2 M}{a_f} \text{ henrys} \quad (14-2-11)$$

$$C = \frac{a_f}{a_x K_s} \text{ farads} \quad (14-2-12)$$

The applied force of 0.5 lb in this example indicates that $a_f = 0.1 \text{ lb/v}$ is a suitable scale factor. From the mechanical value, $\omega_n = 1 \text{ radian/mech. sec}$, $a_t = 10^{-3}$ is chosen so that for the circuit $\omega_n = 1000 \text{ radian/elec. sec}$. Finally, an arbitrary $a_x = 10^6 \text{ ft/coulomb}$ is chosen. Substitution of the scale factors and mechanical parameters into Eqs. 14-2-10, 14-2-11, and 14-2-12 results in

$$R = \frac{(10^6)(0.4)(10^{-3})}{0.1} = 4000 \text{ ohms} \quad (14-2-13)$$

$$L = \frac{(10^6)(0.5)(10^{-6})}{0.1} = 5 \text{ henrys} \quad (14-2-14)$$

$$C = \frac{0.1}{(10^6)(0.5)} = 0.2 \text{ microfarad} \quad (14-2-15)$$

The required driving voltage is found from

$$e_s = \frac{F}{a_f} = \frac{0.5 \text{ lb}}{0.1 \text{ lb/v}} = 5 \text{ v} \quad (14-2-16)$$

The circuit of Fig. 14-2-1a with the above parameter values and driving voltage provides a practical analogy to the mechanical system of this example. For example:

$$q(\tau) = 10^{-6}[1 + 1.09\epsilon^{-400\tau} \sin(917\tau - 113.6^\circ)] \quad (14-2-17)$$

$$\frac{dq(\tau)}{d\tau} = i(\tau) = 1.09 \times 10^{-3} \epsilon^{-400\tau} \sin 917\tau \text{ amp} \quad (14-2-18)$$

$$e_R(\tau) = 4.36 \epsilon^{-400\tau} \sin 917\tau \text{ v} \quad (14-2-19)$$

The use of a_x , a_t , and a_f allows Eqs. 14-2-17, 14-2-18, and 14-2-19 to be converted to

$$x(t) = 1 + 1.09 \epsilon^{-0.4t} \sin(0.917t - 113.6^\circ) \text{ ft} = a_x q(\tau = a_t t) \quad (14-2-20)$$

$$\frac{dx(t)}{dt} = a_x a_t \frac{dq}{d\tau} = 1.09 \epsilon^{-0.4t} \sin 0.917t \text{ ft/mech. sec} \quad (14-2-21)$$

$$f_v \frac{dx(t)}{dt} = a_f e_R = 0.436 \epsilon^{-0.4t} \sin 0.917t \text{ lb} \quad (14-2-22)$$

The other forces, accelerations, voltages, etc. can also be related in a similar manner. Instead of measuring charge directly, the voltage across the capacitance can be measured and then multiplied by C .

Example 14-2-2. In order to extend the force (torque)-voltage analogy to more complicated mechanical problems the system of Fig. 14-2-2 is considered. The mechanical equations are

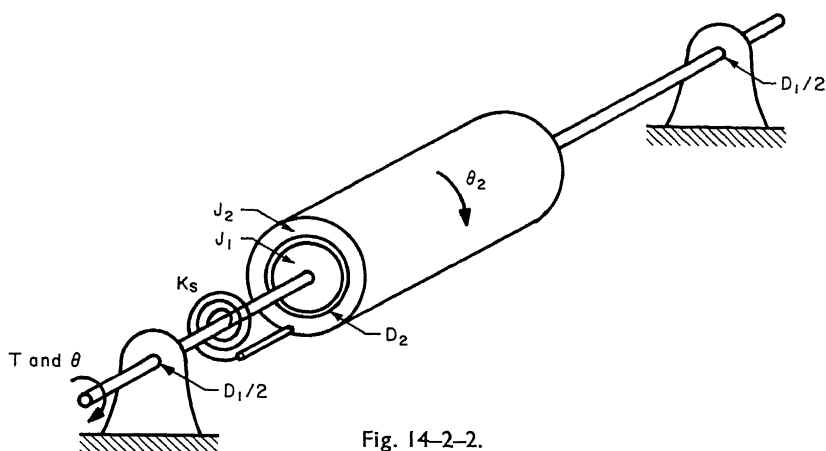


Fig. 14-2-2.

$$J_1 \frac{d^2 \theta_1}{dt^2} + D_1 \frac{d\theta_1}{dt} + D_2 \frac{d(\theta_1 - \theta_2)}{dt} + K_s(\theta_1 - \theta_2) = T \quad (14-2-23)$$

$$J_2 \frac{d^2 \theta_2}{dt^2} + D_2 \frac{d(\theta_2 - \theta_1)}{dt} + K_s(\theta_2 - \theta_1) = 0 \quad (14-2-24)$$

By direct application of the analogies of Table 14-2-1 the electrical equations are

$$L_1 \frac{d^2 q_1}{dt^2} + R_1 \frac{dq_1}{dt} + R_2 \frac{d(q_1 - q_2)}{dt} + \frac{1}{C}(q_1 - q_2) = e_s \quad (14-2-25)$$

$$L_2 \frac{d^2 q_2}{dt^2} + R_2 \frac{d(q_2 - q_1)}{dt} + \frac{1}{C}(q_2 - q_1) = 0 \quad (14-2-26)$$

In constructing the circuit represented by the above equations it is convenient to think of each charge (or current) as circulating around a closed path or loop. A term such as $(q_2 - q_1)/C$ must represent a branch common to two loops. Loops are then blocked out so that all common branches are included. There should be a common circuit branch whenever two position variables are mechanically coupled by some

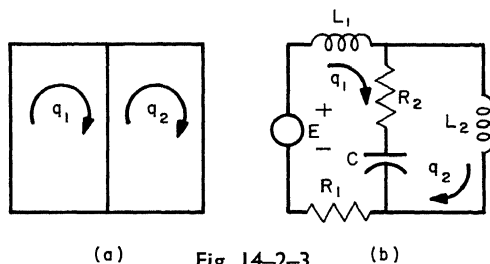


Fig. 14-2-3.

means. Fig. 14-2-3a shows suitable loops for the system being considered. The circuit elements can now be determined from Eqs. 14-2-25 and 14-2-26 and Fig. 14-2-3a. The result is shown in Fig. 14-2-3b. The directions for positive charge flow are arbitrary although it is convenient to have positive charge correspond to positive displacement. The circuit of Fig. 14-2-3b together with suitable changes

in magnitude and time scales as established in Eqs. 14-2-10, 14-2-11, and 14-2-12, can be used to study the behavior of the mechanical system of Fig. 14-2-2. It is interesting to note that the force (torque)-voltage analogy does not result in an inductance common to two loops for most systems. Such an occurrence requires the existence of mutual mass.¹

14-3. The Force (Torque)-Current Mechanical-Electrical Analogy. A second mechanical-electrical analog system is based on the circuit of Fig. 14-3-1 and the mechanical system of Fig. 14-2-1b.

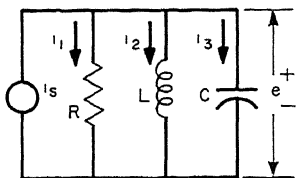


Fig. 14-3-1.

Kirchhoff's current law results in the equation

$$i_1 + i_2 + i_3 = \frac{e}{R} + C \frac{de}{dt} + \frac{1}{L} \int_0^t e dt = i_s \quad (14-3-1)$$

for the circuit of Fig. 14-3-1. In order to establish the required analogy $\int_0^t e dt$ is replaced by a quantity called the magnetic flux linkage ψ . This relation follows from the law of Henry and Faraday which states that

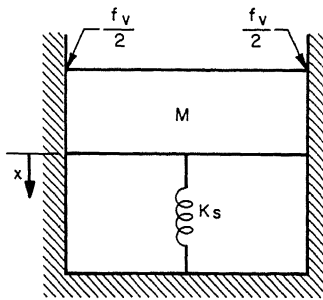
$$e = \frac{d(\text{magnetic flux linkage})}{dt} = \frac{d\psi}{dt} \quad (14-3-2)$$

The substitution of ψ into Eq. 14-3-1 gives

$$C \frac{d^2\psi}{dt^2} + \frac{1}{R} \frac{d\psi}{dt} + \frac{\psi}{L} = i_s \quad (14-3-3)$$

from which the "equality of form" with Eq. 14-3-3 allows Table 14-3-1 to be constructed. The extension of the force (torque)-current analogy to more complicated systems can be based on a circuit node to datum or "ground" voltage for each system velocity. Thus in blocking out an analogous circuit using this scheme, a node is established for each velocity and one more node is added as the reference node. The following examples illustrate the application of the force (torque)-current analogy scheme.

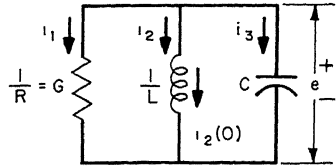
¹ M. F. Gardner and J. L. Barnes, *Transients in Linear Systems*, John Wiley & Sons, New York, 1942, pp. 78-79.



$$M = 2.5 \times 10^{-3} \frac{\text{lb-sec}^2}{\text{in}}$$

$$f_v = 2.5 \times 10^{-2} \frac{\text{lb-sec}}{\text{in.}}$$

$$K_s = 2.5 \times 10^{-1} \text{ lb/in}$$



(a)

Fig. 14-3-2.

(b)

Example 14-3-1. Fig. 14-3-2a shows a mechanical system with mass, viscous friction, and a spring constant. The mass is initially pushed down 6 in. and then released. A practical circuit is to be constructed and scaled so as to accommodate easy electrical study of the system. This problem will involve the initial displacement of the mass which corresponds to initial flux linkage in the circuit. The initial position causes a spring restoring force which can be interpreted as an initial current in the inductance representing the spring. Initial flux linkages are then $\psi(0) = Li_L(0)$. The basic form for the circuit involves two nodes – a reference node and a node whose potential is analogous to the velocity dx/dt . Paths are provided for currents which are analogous to each force in the system. Fig. 14-3-2b shows the resulting circuit and the initial current in L . The mechanical system equation and the circuit equation can now be written as

$$M \frac{d^2 x}{dt^2} + f_v \frac{dx}{dt} + K_s x = 0 \quad (14-3-4)$$

and

$$C \frac{d^2 \psi}{d\tau^2} + \frac{1}{R} \frac{d\psi}{d\tau} + \frac{\psi}{L} = 0 \quad (14-3-5)$$

Direct correspondance of units would have displacement x in inches correspond to flux linkages in webers, force in pounds correspond to current in amperes, velocity in feet/second correspond to volts, and time in seconds in the mechanical system to time in seconds in the electrical circuit. Such a scheme gives $C = 2500 \mu\text{f}$, $R = 40$ ohms, and $L = 4$ h which are rather unsatisfactory values. The method of Ex. 14-2-1 can be used to provide parameter and scale factor relations. For example

$$f_v \frac{dx}{dt} \text{ pounds} = a_f \left[\frac{1}{R} \frac{d\psi(\tau)}{d\tau} \right] \frac{\text{pounds}}{\text{amperes}} \times \text{amperes} \quad (14-3-6)$$

Finally the process of scaling displacement to magnetic flux linkage ($x = a_x \psi$) and mechanical time to electrical time ($\tau = a_t t$) results in the following equations.

$$\frac{1}{G} = R = \frac{a_f}{a_t a_x f_v} \frac{\text{webers}}{\text{amperes-electrical seconds}} \text{ or ohms} \quad (14-3-7)$$

$$L = \frac{a_f}{K_s a_x} \text{ henrys} \quad (14-3-8)$$

$$C = \frac{a_x M a_t^2}{a_f} \text{ farads} \quad (14-3-9)$$

It might be noted that conductance G in this case is the same as R in Eq. 14-2-10. Also, the force-voltage analogy roles of L and C are now interchanged. In fact the force-current analogy provides circuits that are the duals of those resulting from the force-voltage analogy. For the problem of this example the natural frequency ω_n of the mechanical system is found to be 10 radians/sec. This is speeded up arbitrarily to 100 radians/sec in the electrical circuit by using $a_t = 10^{-2}$ elect. sec/mech. sec. A choice of a_f and a_x can be based on the fact that the initial spring force is 1.5 lbs and the maximum mechanical velocity is $x(0)\omega_n$ or 60 in./mech. sec. The velocity is important since it is related to electrical voltage.

$$\text{voltage} = \frac{d\psi}{d\tau} = \frac{1}{a_t a_x} \frac{dx}{dt} \quad (14-3-10)$$

Since 60 would be a suitable voltage, a_x is chosen as 100 in./weber. A quick check of Eqs. 14-3-7, 14-3-8, and 14-3-9 and the initial force, indicates $a_f = 25$ lbs/amp as a choice that results in easily realized values of R , L , and C as well as easily measured currents. For the above scale factor values and mechanical parameter values, $R = 10^3$ ohms, $L = 1$ h, $C = 1$ μ f. The final step requires the value of initial current in L , which can be found from

$$\frac{K_s x(0)}{a_f} = i_L(0) = \frac{\psi(0)}{L} = 0.06 \text{ amp} \quad (14-3-11)$$

or from

$$i_L(0) = \frac{x(0)/a_x}{L} = \frac{\psi(0)}{L} = 0.06 \text{ amp} \quad (14-3-12)$$

It should also be mentioned before leaving this example that the current i_L can be measured instead of ψ , since $\psi = Li_L$. This makes measurements of all required quantities an easy job.

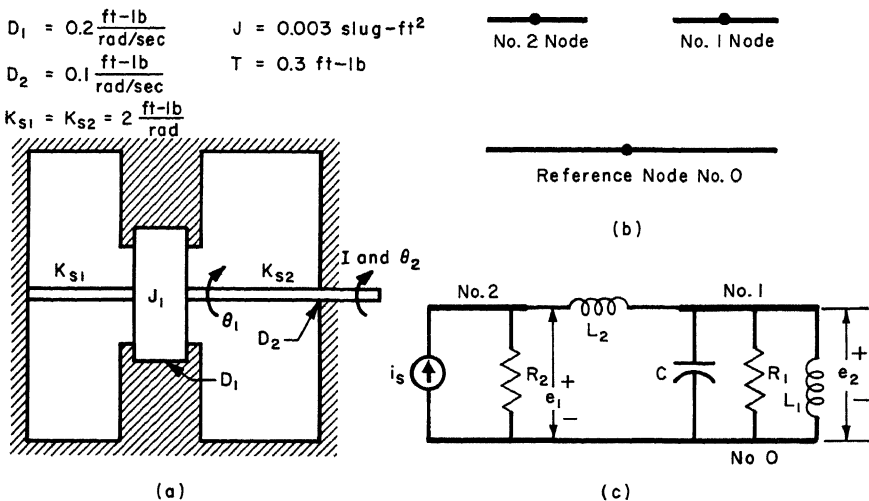


Fig. 14-3-3.

Example 14-3-2. As a more complicated example of the force (torque)-current analogy system, the rotational system of Ex. 4-5-5 is considered. Fig. 14-3-3a repeats the mechanical schematic of Fig. 4-5-4a. The mechanical system equations are

$$J \frac{d^2 \theta_1}{dt^2} + D_1 \frac{d\theta_1}{dt} + K_{s1} \theta_1 + K_{s2} (\theta_1 - \theta_2) = 0 \quad (14-3-13)$$

$$K_{s2} (\theta_2 - \theta_1) + D_2 \frac{d\theta_2}{dt} = T \quad (14-3-14)$$

The equations can be changed to the required electrical equations by use of Table 14-3-1.

TABLE 14-3-1

Electrical		Mechanical			
		Translational		Rotational	
Current	i	Force	F	Torque	T
Magnetic flux linkage = $\int e dt$	ψ	Displacement	x	Angular displacement	θ
Voltage	e	Velocity	dx/dt	Angular velocity	ω
Reciprocal inductance	$1/L$	Spring constant	K_s	Spring constant	K_s
Capacitance	C	Mass	M	Mass moment of inertia	J
Conductance	G $1/R$	Viscous friction coefficient	f_v	Viscous friction coefficient	D

$$C \frac{d^2 \psi_1}{d\tau^2} + \frac{1}{R_1} \frac{d\psi_1}{d\tau} + \frac{\psi_1}{L_1} + \frac{1}{L_2} (\psi_1 - \psi_2) = 0 \quad (14-3-15)$$

$$\frac{1}{L_2} (\psi_2 - \psi_1) + \frac{1}{R_2} \frac{d\psi_2}{d\tau} = i_s \quad (14-3-16)$$

Since two velocity or position variables are needed, the analogous circuit must have three nodes as discussed earlier. Fig. 14-3-3b blocks out a nodal skeleton for such a circuit. Paths are provided for each current that represents a force or for each term in Eqs. 14-3-15 and 14-3-16. Fig. 14-3-3c shows the final circuit together with choices of voltages e_1 and e_2 that are analogous to velocities $d\theta_1/dt$ and $d\theta_2/dt$. If the use of flux linkage variables is confusing in constructing a circuit from Eqs. 14-3-15, and 14-3-16, voltages may be used instead.

$$C \frac{de_1}{d\tau} + \frac{e_1}{R_1} + \frac{1}{L_1} \int_0^t e_1 d\tau + \frac{1}{L_2} \int_0^t (e_1 - e_2) d\tau = 0 \quad (14-3-17)$$

$$\frac{1}{L_2} \int_0^t (e_2 - e_1) d\tau + \frac{e_2}{R_2} = i_s \quad (14-3-18)$$

For the given mechanical parameters the scaling process of Ex. 14-3-1 can now be applied and a practical circuit constructed.

14-4. Hydraulic-Electrical Analogs. The secondary hydraulic parameters of resistance and capacitance of Art. 5-4 in reality form the basis of a system of hydraulic-electrical analogs. In this system hydraulic resistance and capacitance are directly analogous to electrical resistance and capacitance. Table 14-4-1 summarizes this complete system of analogies. The following examples illustrate the application and use of such analogies.

TABLE 14-4-1

Hydraulic	Electrical
Volume of flow	Charge
Volume rate of flow	Current
Head (pressure)	Voltage
Hydraulic resistance	Electrical resistance
Hydraulic capacitance	Electrical capacitance

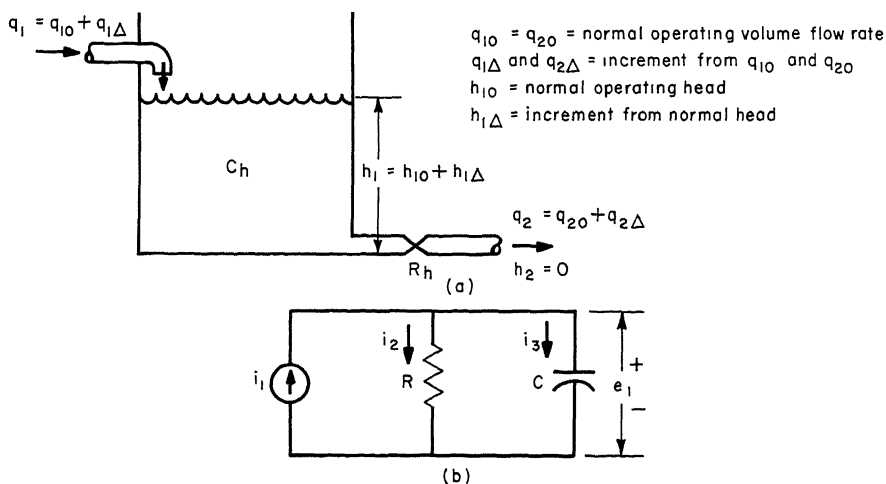


Fig. 14-4-1.

Example 14-4-1. The fluid-handling system of Ex. 5-4-2 serves as a first hydraulic-electrical analog example. Fig. 14-4-1a repeats the system diagram for this example in order to facilitate comparisons. The hydraulic system equations are

$$q_{1\Delta} - q_{2\Delta} = C_h \frac{dh_{1\Delta}}{dt} \quad (14-4-1)$$

$$(h_{1\Delta} - 0) = R_h q_{2\Delta} \quad (14-4-2)$$

$$C_h \frac{dh_{1\Delta}}{dt} + \frac{h_{1\Delta}}{R_h} = q_{1\Delta} \quad (14-4-3)$$

From Table 14-4-1 the electrical equations are

$$i_2 - i_3 = C \frac{de_1}{d\tau} \quad (14-4-4)$$

$$(e_1 - 0) = Ri_2 \quad (14-4-5)$$

$$C \frac{de_1}{d\tau} + \frac{e_1}{R} = i_1 \quad (14-4-6)$$

Fig. 14-4-1b shows the circuit represented by Eqs. 14-4-4, 14-4-5, and 14-4-6. The one voltage and three currents represent the incremental variables in the hydraulic system. Scale factors again play an important role, if an actual circuit is to be constructed and tested; otherwise only general theoretical concepts are aided by the analogy. When scale factors a_q (feet³/second)/ampere, a_h feet/volt, and a_t electrical seconds/hydraulic second are assumed, the electrical parameters are given by

$$C = \frac{a_h C_h a_t}{a_q} \text{ farads} \quad (14-4-7)$$

$$R = \frac{a_q R_h}{a_h} \text{ ohms} \quad (14-4-8)$$

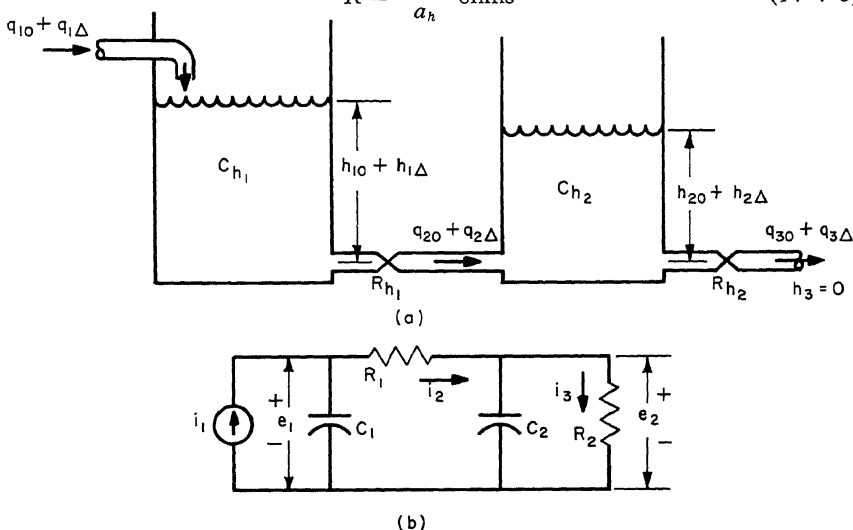


Fig. 14-4-2.

Example 14-4-2. The hydraulic system of Ex. 5-4-3 is shown again in Fig. 14-4-2a. An electrical circuit analog is to be constructed for this hydraulic system. The hydraulic relations are

$$q_{1\Delta} - q_{2\Delta} = C_{h1} \frac{dh_{1\Delta}}{dt} \quad (14-4-9)$$

$$h_{1\Delta} - h_{2\Delta} = R_{h1} q_{2\Delta} \quad (14-4-10)$$

$$q_{2\Delta} - q_{3\Delta} = C_{h2} \frac{dh_{2\Delta}}{dt} \quad (14-4-11)$$

$$h_{2\Delta} - 0 = R_{h2} q_{3\Delta} \quad (14-4-12)$$

The circuit can be constructed without writing the electrical equations by direct use of the analogous quantities. The presence of two head variables requires two voltages for which three nodes are needed. Current paths are provided for each volume flow rate as shown in the completed diagram of Fig. 14-4-2b.

14-5. Electronic Analog Computers—Operational Amplifiers.

As mentioned in Art. 14-1, electronic analog computers can be used to solve the differential equations of physical systems by actually carrying out certain required operations. These computers use voltages as analogs for all dependent system variables and time as the independent variable. The basic building block of such analog computers is a high-gain d-c

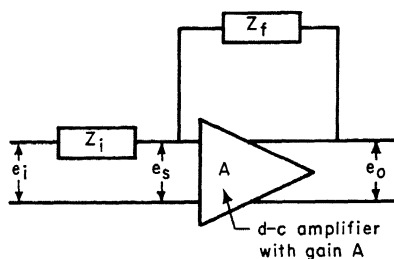


Fig. 14-5-1.

amplifier with various feedback impedance elements. Fig. 14-5-1 shows a schematic representation of an operational amplifier. The equations which define the operation of this circuit are

$$\frac{e_i - e_g}{Z_i} = \frac{e_g - e_o}{Z_f} \quad (14-5-1)$$

$$e_o = A e_g \quad (14-5-2)$$

The first of these equations assumes that no current flows into the amplifier A . If e_g is eliminated by substituting Eq. 14-5-2 into Eq. 14-5-1,

$$\frac{e_o}{e_i} = -\frac{Z_f}{Z_i} \left[\frac{1}{1 - \frac{1}{A} \left(1 + \frac{Z_f}{Z_i} \right)} \right] \quad (14-5-3)$$

In general A is much greater than unity and Eq. 14-5-3 can be approximated by

$$\frac{E_o(s)}{E_i(s)} \approx \frac{-Z_f(s)}{Z_i(s)} \quad (14-5-4)$$

which is of basic importance in analog computer work. Several useful operations are indicated in Fig. 14-5-2 together with standard symbols for the internal amplifier and for the over-all operational amplifier. These few units can be used to solve a great many problems. More complete

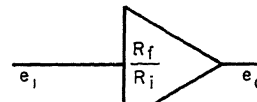
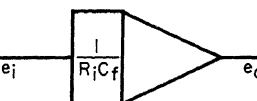
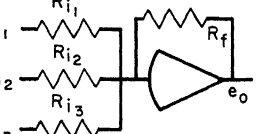
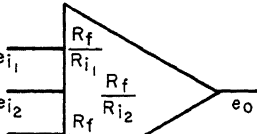
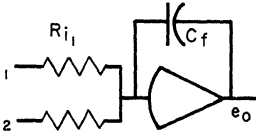
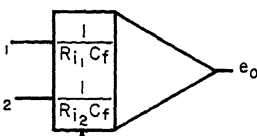
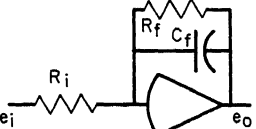
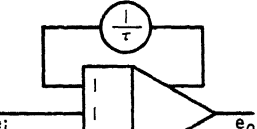
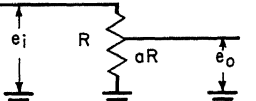
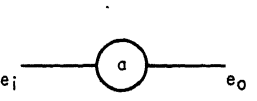
 Constant multiplier	$\frac{E_o(s)}{E_i(s)} = -\frac{R_f}{R_i}$	
 Integrator	$\frac{E_o(s)}{E_i(s)} = -\frac{1}{R_i C_f s}$ $e_o(\tau) = -\frac{1}{R_i C_f} \int_0^\tau e_i d\tau$	
 Adder	$e_o = -\left[\frac{R_f}{R_{i1}} e_{i1} + \frac{R_f}{R_{i2}} e_{i2} + \frac{R_f}{R_{i3}} e_{i3} \right]$	
 Summing integrator	$E_o(s) = -\frac{E_{i1}(s)}{R_{i1} C_f s} - \frac{E_{i2}(s)}{R_{i2} C_f s}$	
 Lag unit	$\frac{E_o(s)}{E_i(s)} = -\frac{R_f}{R_i} \frac{1}{1 + R_f C_f s}$ or $\frac{E_o(s)}{E_i(s)} = -\frac{\tau}{1 + \tau s}$	
 Coefficient potentiometer	$\frac{E_o}{E_i} = a < 1$	

Fig. 14-5-2.

tables of operations can be found in the literature of the subject. It might be mentioned at this point that interchanging the Z_i and Z_f networks simply inverts the operation performed by the amplifier.

Since no differentiator is listed in Fig. 14-5-2, it might appear that differential equations could not be solved; that such is not the case is shown in Art. 14-6. The operation of differentiation is not ordinarily used, since it amplifies noise because of its ever-increasing gain vs. frequency curve as obtained from $G(j\omega) = j\omega$. Normally, signal voltages must be kept large enough to mask inherent noise. There are also limitations to the operation of the unit shown in Fig. 14-5-2. The output of all amplifier units is limited to voltages between ± 100 volts or ± 50 volts

Example 14-6-1. A computer circuit for solving the classic equation

$$\frac{d^2x}{dt^2} + a_1 \frac{dx}{dt} + a_0x = f(t) \quad (14-6-1)$$

is to be drawn. At this point no attention is given to the limitations of an actual electronic analog computer. In fact a computer that would use the actual dependent variable x and independent variable t is used. This temporarily bypasses the scaling problem. A standard procedure commonly used to eliminate the need for differentiators is to solve the equations for the highest derivative.

$$\frac{d^2x}{dt^2} = f(t) - a_1 \frac{dx}{dt} - a_0x \quad (14-6-2)$$

The quantity d^2x/dt^2 is assumed to be a known signal that can be integrated to give $-dx/dt$ as shown in Fig. 14-6-1a. Multiplication by a constant a_1 is also carried out in the first integrator as well as insertion of the proper initial value of dx/dt . The multiplication by a_1 could be done by a coefficient pot if a_1 is less than unity. Fig. 14-6-1a also shows a second integrator that converts $-a_1 dx/dt$ to a_0x . Another coefficient pot could be used here. According to Eq. 14-6-2, d^2x/dt^2 can be formed by adding $f(t)$, $-a_1 dx/dt$, and $-a_0x$. This operation is indicated in Fig. 14-6-1b in which the sign reversal of the operational amplifier must always be considered. The signal a_0x is directly available from the diagram of Fig. 14-6-1a from which $a_1 dx/dt$ can also be obtained by use of a sign changing operational amplifier with unity gain. Presumably the function $f(t)$ can be externally generated so that interconnection of all units results in the computer diagram of Fig. 14-6-1c for the solution of Eq. 14-6-1. An alternate scheme is shown in Fig. 14-6-1d which assumes that a_1 and a_0 are less than unity and uses two coefficient pots. A rather significant point in both Figs. 14-6-1c and d is that an odd number of operational units is present in tracing in the direction of signal flow around each closed loop. In some cases it may not be necessary to have d^2x/dt^2 actually available as a signal. A summing integrator can then be used to eliminate the adder unit. The result of using this simplification is shown in Fig. 14-6-2. The formation of $-dx/dt$ as shown can be checked as follows

$$-\frac{dx}{dt} = - \int_0^t \frac{d^2x}{dt^2} dt = - \int_0^t [f(t) - a_0x - a_1 \frac{dx}{dt}] dt \quad (14-6-3)$$

Again an odd number of operational amplifiers is included in each forward loop.

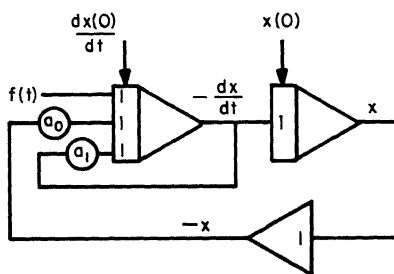


Fig. 14-6-2.

Example 14-6-2. A computer diagram is to be constructed to solve the equations of the mechanical system of Ex. 14-2-2. Zero initial conditions are assumed. The system Eqs. 14-2-22 and 14-2-23 are solved for the highest derivatives as follows

$$\frac{d^2\theta_1}{dt^2} = \frac{T}{J} - \frac{(D_1 + D_2)}{J} \frac{d\theta_1}{dt} + \frac{D_2}{J} \frac{d\theta_2}{dt} - \frac{K}{J} \theta_1 + \frac{K_s}{J} \theta_2 \quad (14-6-4)$$

$$\frac{d^2\theta_2}{dt^2} = -\frac{D_2}{J} \frac{d\theta_2}{dt} + \frac{D_2}{J} \frac{d\theta_1}{dt} - \frac{K_s}{J} \theta_2 + \frac{K_s}{J} \theta_1 \quad (14-6-5)$$

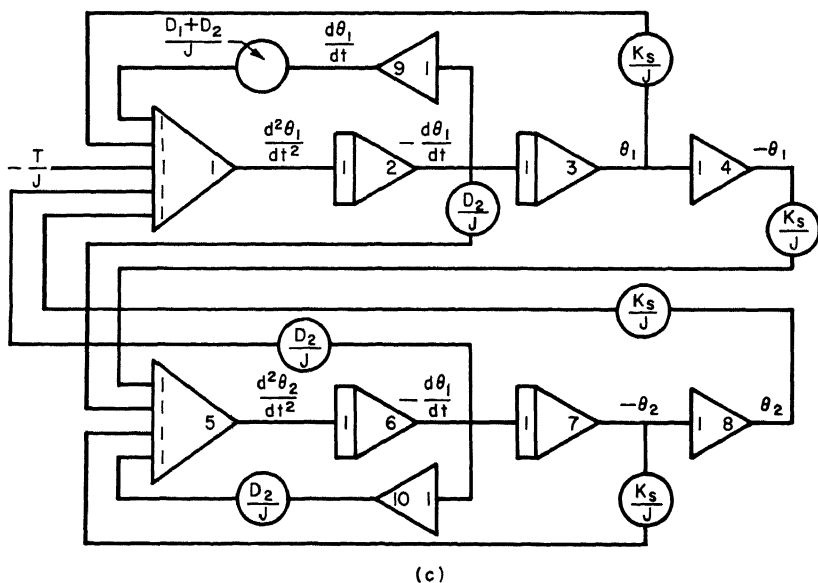
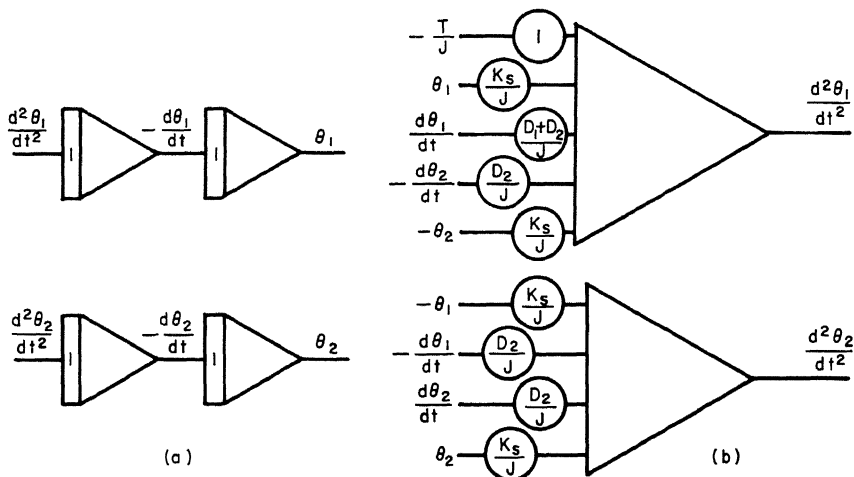


Fig. 14-6-3.

Both second derivatives are now assumed to be known signals; thus successive integrations can be used to obtain the first derivatives and then the dependent variables as shown in Fig. 14-6-3a. The second derivatives are formed by adding the terms indicated on the right side of Eqs. 14-6-4 and 14-6-5 as shown in Fig. 14-6-3b. Interconnection of the θ_1 and θ_2 diagrams plus the use of several sign changes results in the final diagram of Fig. 14-6-3c. The coefficient pots and the adder gains can be adjusted as required in case a coefficient is greater than unity. In most cases the coefficients must be split up and distributed around the system in order to obtain suitable operating signal levels. Frequently the actual interconnecting lines are omitted and some identifying designation is used to indicate connections. This makes the diagram appear simpler.

The above two general examples illustrate the equation solving technique used in putting problems on an analog computer. A second approach is direct simulation which minimizes the number of components required but makes fewer signals available for study in some cases. The next two examples illustrate this second approach for some simple closed-loop control systems.

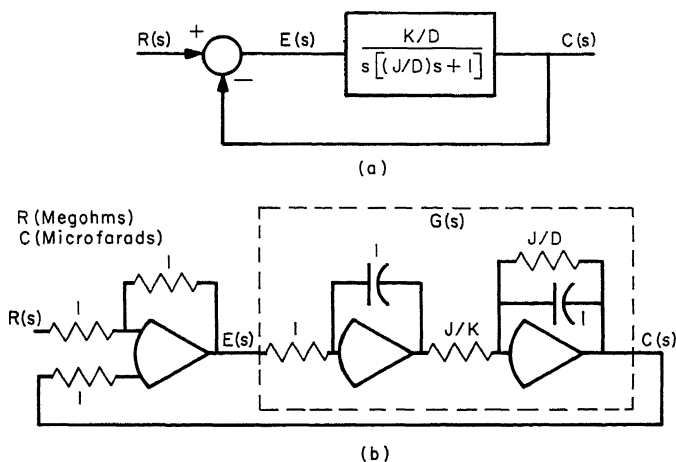
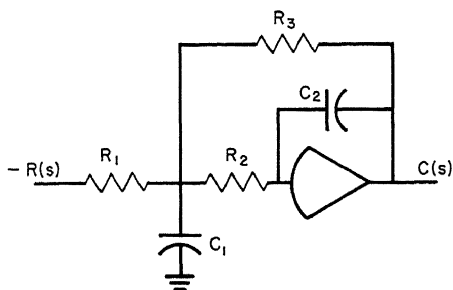


Fig. 14-6-4.

Example 14-6-3. The basic second-order unity feedback control system of Art. 6-8 is to be studied by using an analog computer. Again scaling of physical quantities to computer voltages is not considered at this point. Fig. 14-6-4a shows the servo block diagram. An analog computer block diagram can be constructed by breaking the transfer function $G(s)$ into two blocks—one an integration, $1/s$, the other a lag operation, $(K/D)/[(J/D)s + 1]$. These two operations can be easily performed by operational amplifiers. An adder can be used as the summing point. The results of putting the computer units together is shown in Fig. 14-6-4b. It should be noted that the summing point performs the operation.

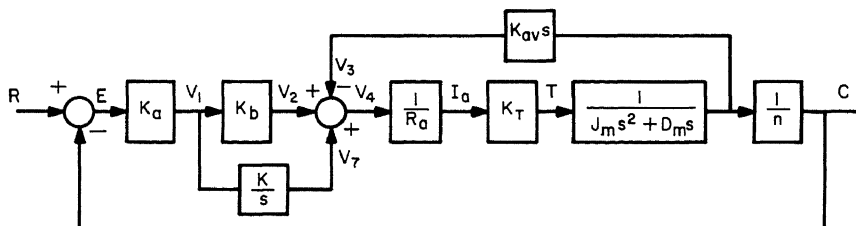
$$E(s) = -[-R(s) + C(s)] \quad (14-6-6)$$

Again, the gain terms can be redistributed to provide more suitable signal levels. The similarity of form between the servo block diagram and the computer diagram indicates that the computer more or less simulates the actual system. Considerably more involved transfer functions can be set up on an analog computer as is demonstrated by the fact that the whole system of Fig. 14-6-4a can be represented by the computer circuit of Fig. 14-6-5, if the initial conditions are zero.

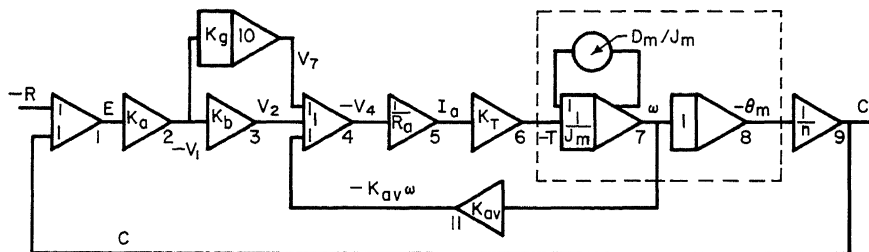


$$\frac{C(s)}{R(s)} = -\frac{R_3}{R_1} \frac{1}{(R_2 R_3 C_1 C_2) S^2 + C_2 (R_2 + R_3 + \frac{R_2 R_3}{R_1}) S + 1}$$

Fig. 14-6-5.



(a)



(b)

Fig. 14-6-6.

Example 14-6-4. A second example of the use of a computer circuit based on a physical system block diagram is the system of Fig. 14-6-6a which is studied in Art. 6-18. A computer diagram preserving all variables but having zero initial conditions is to be constructed. The form of the block diagram can be retained almost intact. However, in order to eliminate the need for differentiating the motor position variable to obtain the counter emf term $K_A V_s \theta_m$, the torque-to-position

transfer block is split into two parts as shown. The lag network unit no. 7 is slightly modified to give the transfer function $1/(J_ms + D_m)$. Undoubtedly in any practical problem a number of the amplifiers could be eliminated by combining gains.

The simple basic uses of electronic analog computers have been indicated with no discussion as to how to actually mechanize a problem. Inevitably a problem must have physical dependent and sometimes independent variables scaled to analogous machine voltages, and independent variables scaled to machine time in actual seconds. This process must take into account the limitations on computer equipment listed in Art. 14-5 as well as limitations on recording or display equipment. The following articles discuss the above problem.

14-7. Analog Computers—Time Scaling. In most physical problems of interest in control systems, the independent variable is time. Since a wide range of system response times is encountered, and since most computers have rather limited computing times a change in time scale must generally be made in putting a problem on an analog computer. In many analog computers the maximum signal frequency should not exceed 10 cps especially when servo multiplying units are used. A time scale change is accomplished by substituting actual computer running time, τ , into the system equations according to a relation such as

$$\tau = a_t t \text{ computer seconds} \quad (14-7-1)$$

The scale factor a_t is the number of computer seconds equal to a second in the original problem. Thus if a_t is 10, then 1 second in the original problem now takes 10 seconds on the computer. If a_t is greater than unity, the problem is slowed down whereas if a_t is less than unity the problem is speeded up. From Eq. 14-7-1 it follows that

$$\frac{d(\quad)}{dt} = a_t \frac{d(\quad)}{d\tau} \quad (14-7-2)$$

$$s(\quad) = a_t s'(\quad) \quad (14-7-3)$$

Where s and s' are the Laplace variables for different time bases. It should be carefully noted that slowing a problem down decreases the value of derivatives when taken with respect to the new time variable. The following examples illustrate a change in the time base.

Example 14-7-1. The simple mass-friction-spring mechanical system with

$$3 \frac{d^2 x}{dt^2} + 90 \frac{dx}{dt} + 900x = 90 \sin 62.8t \text{ lbs} \quad (14-7-4)$$

is to be slowed down for analog computer study. A preliminary analysis indicates a natural undamped angular velocity $\omega_n = 17.32$ radians/sec. A slowdown by a factor of two might be suitable. Thus $\tau = 2t$ is used to give

$$12 \frac{d^2 x}{d\tau^2} + 180 \frac{dx}{d\tau} + 900x = 90 \sin 31.4\tau \text{ lbs} \quad (14-7-5)$$

In general the time scale change can be associated with the integrators of an analog computer system as considered in the next example.

Example 14-7-2. The generalized form of the mechanical system equation of Ex. 14-7-1 is reconsidered as follows:

$$a_t^2 M \frac{d^2 x}{d\tau^2} + a_t f_v \frac{dx}{d\tau} + K_s x = f(\tau) \quad (14-7-6)$$

$$a_t^2 \frac{d^2 x}{d\tau^2} = \frac{f(\tau)}{M} - \frac{a_t f_v}{M} \frac{dx}{d\tau} - \frac{K_s x}{M} \quad (14-7-7)$$

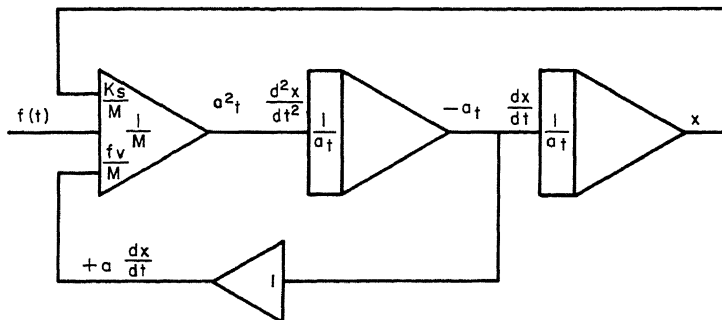


Fig. 14-7-1.

The latter equation can be set up as indicated in Fig. 14-7-1. The time scale factor a_t appears only as a gain term for the integrators. Generally a change in a_t is accomplished by multiplying the integrator feedback capacitors by a_t . This also follows from the fact that only capacitance is associated with derivatives of voltage in the operational amplifiers. Since the

$$C \frac{d(\quad)}{dt} = (a_t C) \frac{d(\quad)}{d\tau} \quad (14-7-8)$$

gain around a computer loop can be redistributed, the time scale change factor can also be handled in other ways.

Example 14-7-3. As a final time-scale change example, the control system of Ex. 14-6-3 is to have its time scale changed by a_t . This can be accomplished in Fig. 14-6-4 by replacing s by $(a_t s')$ and by changing the integrator and lag unit capacitors from one microfarad to a_t microfarads. These steps can be checked as follows. The system differential equation is

$$\frac{J}{D} \frac{d^2 c}{dt^2} + \frac{dc}{dt} = \frac{K}{D} e(t) \quad (14-7-9)$$

which can be changed to

$$a_t^2 \frac{J}{D} \frac{d^2 c}{d\tau^2} + a_t \frac{dc}{d\tau} = \frac{K}{D} e(\tau) \quad (14-7-10)$$

The transform with respect to the τ base is

$$\frac{C(s')}{E(s')} = \frac{K/D}{(a_t s') \left[\frac{J}{D} (a_t s') + 1 \right]} \quad (14-7-11)$$

which also results from the substitution indicated above. The change in the computer diagram follows from the equations for the integrator and lag unit.

$$\frac{1}{R_i C_f s'} = \frac{1}{a_i s'} \quad ; \quad R_i = 10^6, C_f = a_i \times 10^{-6} \quad (14-7-12)$$

$$\frac{R_f/R_i}{R_f C_f s' + 1} = \frac{K/D}{(J/D)a_i s' + 1} \quad ; \quad R_f = \frac{J}{D} \times 10^6$$

$$R_i = \frac{J}{K} \times 10^6$$

$$C_f = a_i \times 10^{-6} \quad (14-7-13)$$

14-8. Analog Computers—Amplitude Scaling. Following suitable time scale changes in putting problems on an analog computer, attention must be devoted to scaling of amplitudes. The computer manipulates voltages and therefore scale factors such as pounds/(unit of computer voltage) must be chosen so that the voltages at each point in the computer do not exceed the amplifier limits. Also voltages must not be so small that stray pickup and noise is a problem. The first step in this procedure requires rough determination of the maximum expected amplitudes of the actual physical quantities. This information may come from a knowledge of the actual system, from a preliminary analysis of the system equations, or from a pure guess. In many cases estimates are made by neglecting the damping in a system. Thus if a system variable $y_{\max}$ is known, then $[dy/dt]_{\max} \approx \omega_n y_{\max}$ and $\int_0^t y dt_{\max} \approx y_{\max}/\omega_n$ etc. These values are obtained from the undamped equation

$$y = y_{\max} \sin(\omega_n t + \varphi) \quad (14-8-1)$$

and must be used with caution since in some cases an oscillation about a steady value exists.

Example 14-8-1. Estimates are to be obtained from the system Eqs. 14-8-2 and 14-8-3 for maximum expected values of all variables, derivatives, and integrals. Zero initial conditions are assumed. As a first step the coupling between the variables

$$100i_1 + 0.1 \frac{di_1}{dt} + 10^7 \int_0^t i_1 dt - 0.1 \frac{di_2}{dt} = 10 \quad (14-8-2)$$

$$-0.1 \frac{di_1}{dt} + 0.3 \frac{di_2}{dt} + 10^7 \int_0^t i_2 dt = 0 \quad (14-8-3)$$

i_1 and i_2 are ignored in Eq. 14-8-2. The damping term $100i_1$ is also neglected. These steps allow natural undamped and uncoupled frequencies to be determined by analogy to the basic expression $s^2 + \omega_n^2$ or $s + \omega_n/s$:

$$\omega_{n_1} \approx \sqrt{\frac{10^7}{0.1}} = 10^4 \text{ radians/sec} \quad (14-8-4)$$

$$\omega_{n_2} \approx \sqrt{\frac{10^7}{0.3}} = 5.77 \times 10^3 \text{ radians/sec} \quad (14-8-5)$$

In order to determine a rough value for $i_{1\max}$ the equation

$$I_1(s) \approx \frac{10}{s^2 + \omega_{n_1}^2} \quad (14-8-6)$$

is used to obtain $i_{1\max} \approx 0.01$ amp. From this value and ω_{n_1} , $[di_1/dt]_{\max} \approx 100$ amp/sec and $[\int_0^t i_1 dt]_{\max} \approx 10^{-6}$ coulomb. The second equation is used now with the maximum value of di_1/dt to give

$$\left[\frac{di_2}{dt} \right]_{\max} \approx \frac{0.1}{0.3} (100) = 33 \text{ amp/sec} \quad (14-8-7)$$

$$\left[\int_0^t i_2 dt \right]_{\max} \approx \frac{33}{\omega_{n_2}^2} \approx 10^{-6} \text{ coulomb} \quad (14-8-8)$$

The actual values for the desired quantities excluding damping but including coupling are

$$i_{1\max} \approx 0.019 \text{ amp} ; \quad i_{2\max} \approx 0.0065 \text{ amp}$$

$$\left. \frac{di_1}{dt} \right|_{\max} \approx 188 \text{ amp/sec} ; \quad \left. \frac{di_2}{dt} \right|_{\max} \approx 66 \text{ amp/sec} \quad (14-8-9)$$

$$\int_0^t i_1 dt|_{\max} \approx 2.2 \times 10^{-6} \text{ coulomb} ; \quad \int_0^t i_2 dt|_{\max} \approx 3.8 \times 10^{-6} \text{ coulomb.}$$

The estimates check within reason of the above values and are easy to obtain.

After initial estimates of magnitudes are obtained, suitable amplitude scale factors can be established between computer voltages and system variables. In order to minimize confusion at this point the time scale change is carried out by substitution of $\tau = a_t t$ into the system equations before magnitudes are scaled. The following scheme of amplitude scaling attempts to retain an easily visualized correspondence between machine voltages and system variables.² Fig. 14-8-1

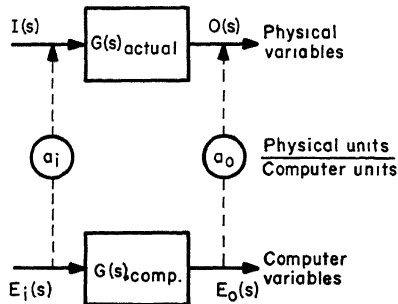


Fig. 14-8-1.

illustrates an operation in both a physical block diagram and the corresponding computer diagram. The factors a_i and a_o are the amplitude scale factors in physical units/computer volt or unit. For example, with a computer volt termed a computer unit a_i might be (feet/sec)/computer unit while a_o is feet/computer unit. The scale factors are generally chosen on the basis of the maximum values of the physical

² L. Jubin Lane, "A Method of Scaling and Checking Computer Circuits," *Trans. AIEE* 77, Part II (May, 1958): 67-70.

variables and the fact that computer voltages should run between ± 10 and ± 90 volts at the output of the amplifiers of most computers. For some computers the maximum voltage should not exceed ± 50 volts. In order to determine the operation $G(s)_{\text{comp}}$ for given scale factors a_i and a_o and given $G(s)_{\text{actual}}$ the following steps are used.

$$\begin{aligned} O(s) &= G(s)_{\text{actual}} I(s) \\ E_o(s) &= G(s)_{\text{comp}} E_i(s) \\ a_i E_i(s) &= I(s) \\ a_o E_o(s) &= O(s) \end{aligned} \quad (14-8-10)$$

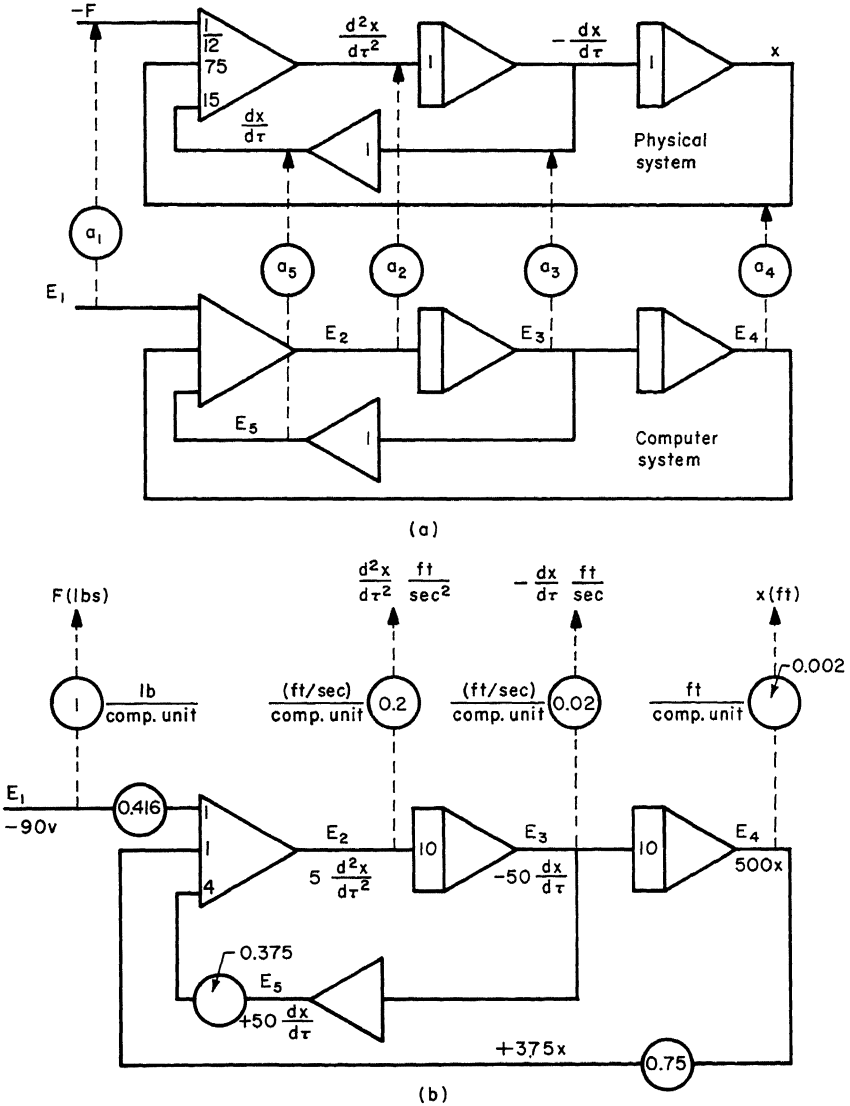


Fig. 14-8-2.

From Eq. 14-8-10 the relation

$$G(s)_{\text{comp}} = \frac{a_i}{a_o} G(s)_{\text{actual}} \quad (14-8-11)$$

can be obtained. Such an equation should be set up for each block where scale factors change as is shown in the examples that follow. If operational amplifiers with fixed input and feedback impedances are used, some redistributions of gain and scale factors plus use of coefficient potentiometers may have to be accomplished.

Example 14-8-2. The mechanical system of Ex. 14-7-1 is to be amplitude scaled for solution on an analog computer. Only gains of 1, 4, and 10 are available on the operational amplifiers. The slow time scale of Eq. 14-7-5 is to be used, but a new step function driving force of 90 lbs is applied instead of the sinusoidal force. Zero initial conditions are assumed. The equation to scale is

$$\frac{d^2x}{d\tau^2} = \frac{F}{12} - 15\frac{dx}{d\tau} - 75x \quad ; \quad F = 90 \text{ lbs} \quad (14-8-12)$$

Fig. 14-8-2a shows the correspondence between physical system diagram and computer diagram. In order to establish numerical values for the scale factors a_1 through a_5 estimates are made of x_{max} , $(dx/d\tau)_{\text{max}}$, and $(\int_0^t id\tau)_{\text{max}}$. From Eq. 14-8-12 ω_{n_τ} is 8.67 radians/sec and $\zeta = 0.866$. Also the final value of x is 0.1 ft. Thus

TABLE 14-8-1

Quantity	Estimated maximum value	Scale factor
x	0.1 ft	$a_4 = 0.002 \text{ ft/comp. unit}$
$dx/d\tau$	0.867 ft/sec	$a_3 = a_5 = 0.02 \text{ (ft/sec)/comp. unit}$
$d^2x/d\tau^2$	7.5 ft/sec ²	$a_2 = 0.2 \text{ (ft/sec}^2\text{)/comp. unit}$
F	90 lbs	$a_1 = 1.0 \text{ lb/comp. unit}$

reasonable estimates of the required quantities are as shown in Table 14-8-1 together with scale factor values that give operational amplifier output voltages in the range of 50 v. The computer gains are found from Eq. 14-8-11 as follows.

$$\begin{aligned} \frac{E_3}{E_2} &= G_{1c-2c} = \left(\frac{a_2}{a_3} \right) \left(\frac{1}{s} \right) = \frac{0.2}{0.02} \left(\frac{1}{s} \right) = \frac{1}{R_i C_f s} \quad ; \quad \frac{1}{R_i C_f} = 10 \text{ sec}^{-1} \\ \frac{E_4}{E_3} &= \left(\frac{a_3}{a_4} \right) \left(\frac{1}{s} \right) = \left(\frac{0.02}{0.002} \right) \left(\frac{1}{s} \right) \quad ; \quad \frac{1}{R_i C_f} = 10 \text{ sec}^{-1} \\ \frac{E_4}{E_2} &= \left(\frac{a_4}{a_2} \right) (75) = \left(\frac{0.002}{0.2} \right) (75) = 0.75 \\ \frac{E_2}{E_5} &= \left(\frac{a_5}{a_2} \right) (15) = \left(\frac{0.02}{0.2} \right) (15) = 1.5 \\ \frac{E_2}{E_1} &= \left(\frac{a_1}{a_2} \right) \left(\frac{1}{12} \right) = \left(\frac{1}{0.2} \right) \left(\frac{1}{12} \right) = 0.416 \end{aligned} \quad (14-8-13)$$

In order to use the standard amplifier gains, coefficient pots are included in the final system as shown in Fig. 14-8-2b. Several new scale factors could now be used following the coefficient pots but these are not of much importance. In reality, on a unit to unit basis the computer is generating multiplies such as $5d^2x/d\tau^2$ of the physical system quantities as shown in Fig. 14-8-2b. If initial conditions were present, they could be introduced by adders following each integrator or by charging integrator capacitors to the correct initial voltages. For example, voltages numerically equal to $-50dx(0)/d\tau$ and $500x(0)$ would be added to voltage E_2 and E_3 respectively. There are several checks on the scaling process. Each equation in set of Eq. 14-8-13 should check dimensionally. For example,

$$\frac{E_2}{E_4} = \left(\frac{a_4}{a_2}\right) \left(\frac{K_s}{M}\right) \left(\frac{\text{ft/comp. unit}}{(\text{ft/sec}^2)/\text{comp. unit}}\right) \left(\frac{\text{lbs/ft}}{\text{lbs-sec}^2/\text{ft}}\right) \rightarrow \text{Dimensionless} \quad (14-8-14)$$

as required for a ratio of resistances in a coefficient pot and in an adder. A second check involves the product of loop gains in the actual system diagram and in the computer diagram. These loop gains must be identical as can be seen in Fig. 14-8-3.

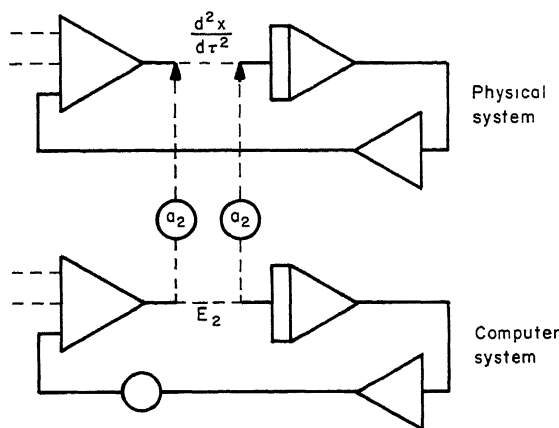


Fig. 14-8-3.

If an equation of the form of 14-8-11 is written with the whole loop $G(s)$ values, the scale factor cancels and the physical system loop $G(s)_{\text{actual}}$ must equal the computer loop $G(s)_{\text{comp}}$. The student should check the two loops in Fig. 14-8-2 using this scheme. If for some reason the computer voltages are too high and overload the amplifier for the scale factors used, the driving voltage and initial condition voltages can be decreased by the same factor until all overloads disappear (presuming that scaling is not too far off). All amplitude scale factors must then be multiplied by the ratio of the original driving voltage to the new driving voltage. The changes in scale factors cancel when setting up the scaling Eq. 14-8-11 so that no computer blocks change. For example, if E_1 in Fig. 14-8-2 were changed to -45 v, A_1 would become 2 lbs/comp. unit, A_4 would be 0.004 ft/comp. unit, etc. On a unit to unit basis, E_2 is now only $2.5 d^2x/d\tau^2$, E_3 is $-25 dx/d\tau$, etc. This scheme is restricted to linear systems.

Example 14-8-3. As an example of scaling a control system problem, the simple mechanical positioning system of Ex. 14-6-3 is reconsidered. Fig. 14-8-4a shows the system block diagram while Fig. 14-8-4b gives a physical variable computer

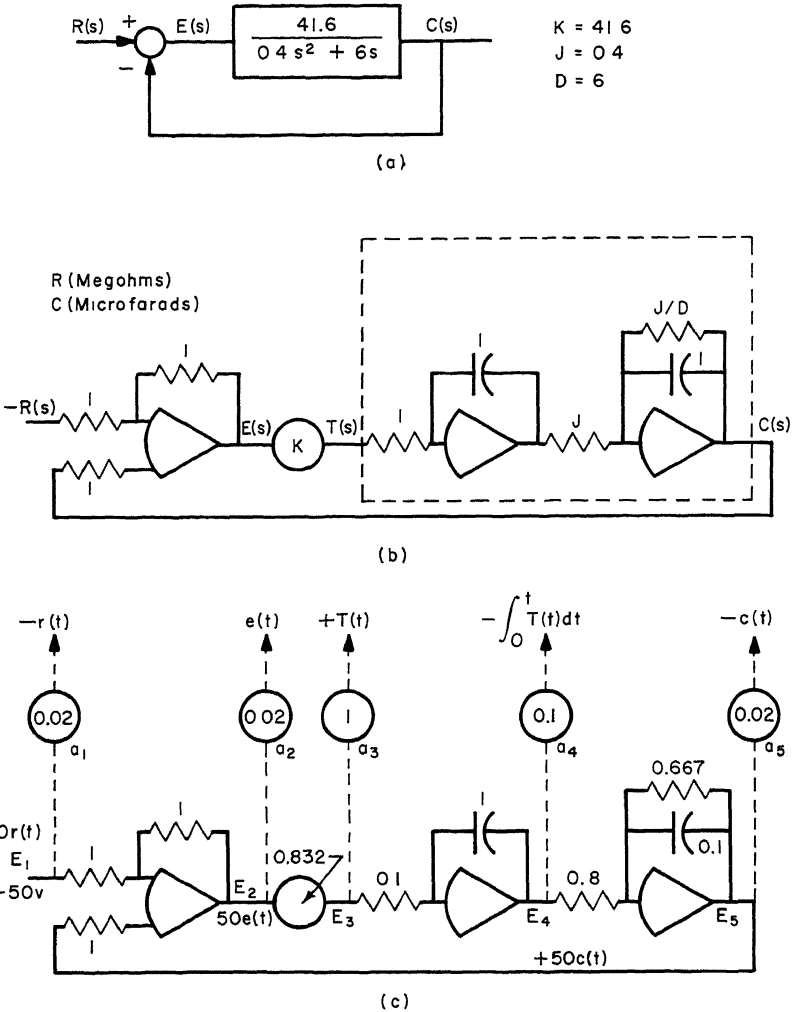


Fig. 14-8-4.

TABLE 14-8-2

Physical quantity	Scale factor
$e(t)_{\max} \approx 1 \text{ rad}$	$a_2 = 0.02 \text{ rad/comp. unit}$
$c(t)_{\max} \approx 1 \text{ rad}$	$a_5 = 0.02 \text{ rad/comp. unit}$
$T(t)_{\max} \approx 40 \text{ in.-lb}$	$a_3 = 1 \text{ (in.-lb)/comp. unit}$
$\int_0^t T(t) dt_{\max} \approx 5 \text{ in.-lb-sec}$	$a_4 = 0.1 \text{ (in.-lb-sec)/comp. unit}$
$r(t) = 1 \text{ rad}$	$a_1 = 0.02 \text{ rad/comp. unit}$

diagram. Preliminary estimates of the various physical quantities for a unit-step input are found from $\omega_n = \sqrt{41.6/0.4} = 10.2$ radians/sec, $\zeta = 0.735$, and the final-value theorem. Table 14-8-2 presents the needed values plus a choice of scale factors that give reasonable voltages. Only the lag unit scaling is considered in detail.

$$\frac{R_f/R_i}{R_f C_f s + 1} = \left(\frac{a_2}{a_3} \right) \left(\frac{1/6}{(0.4/6)s + 1} \right) = \frac{5/6}{(1/15)s + 1} \quad (14-8-15)$$

For $C_f = 0.1 \mu\text{f}$, $R_f = 0.667$ megohm and $R_i = 0.8$ megohm. The checking techniques and rescaling scheme of Ex. 14-8-3 can also be applied here.

Only simple linear computer problems have been considered thus far. In fact many of the specialized problems encountered in linear systems on various computers have not been discussed because of space limitations and the desire to deal with basic considerations. For more detailed analog computer study the student is referred to the literature.^{3,4} It would be unfair, however, if at least a glimpse of some of the nonlinear operations performed on analog computers were not included. The final article of this chapter undertakes such a discussion.

14-9. Nonlinear Operations and Characteristics. The operation of multiplication is generally termed a nonlinear operation although multiplication by the independent variable may be linear. In computer operation some scheme for multiplication must be available. One common device performing this operation is an actual instrument servomechanism

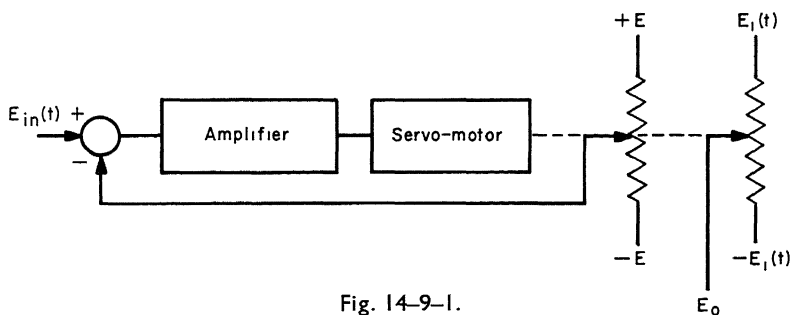


Fig. 14-9-1.

as shown in Fig. 14-9-1. A voltage E_{in} into the multiplier causes the servomotor to drive until the error is zero at which time the voltage fed back, βE , is equal to E_{in} . Thus the fractional pickoff on the potentiometers is $\beta = E_{in}/E$. The output voltage E_o is then βE_1 or

$$E_o = \frac{E_{in} E_1}{E} \quad (14-9-1)$$

³ Clarence L. Johnson, *Analog Computer Techniques*, McGraw-Hill Book Co., Inc., New York, 1956.

⁴ G. A. Korn and T. M. Korn, *Electronic Analog Computers*, McGraw-Hill Book Co., Inc., New York, 1952.

and E_1 has been multiplied by E_{1n} . The operation of this type of multiplier is slow because of the servomotor and mechanical components. The time scale on a problem must be chosen so that the motor can follow the variations in $E_{1n}(\tau)$ with negligible error. All electronic multipliers are available which eliminate this speed restriction, but discussion of these units is beyond the scope of this text.

Physical systems invariably possess some additional kinds of nonlinearities that must be considered in precise studies. A saturation or limiting type of phenomenon is quite commonly encountered. Fig.

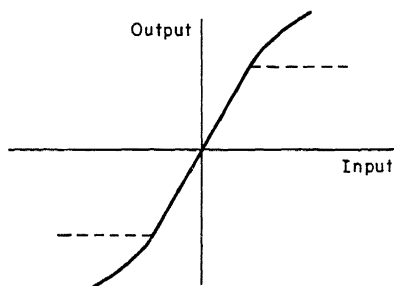


Fig. 14-9-2.

14-9-2 illustrates the input-output graphical relationship for this type of nonlinearity. This graph simply illustrates the fact that the output of a component in a system does not increase indefinitely as the input is increased. A typical example of this phenomenon is magnetic saturation.

Analog computers can be made to include this nonlinearity by the use of vacuum or crystal diodes. Care must be exercised in setting up and scaling a computer circuit with this nonlinearity so that the limiting or saturation levels in the computer correspond correctly to those in physical problem.

A second nonlinear type of behavior results in a so-called dead zone as shown in Fig. 14-9-3. In this case, no output appears until the input

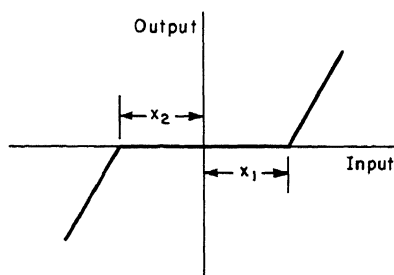


Fig. 14-9-3.

exceeds x_1 or x_2 . Motors with static friction exhibit this type of phenomenon since the input voltage must reach a certain level before enough torque is developed to overcome the stiction torque. Simple diode circuits can be used with analog computer operational amplifiers to simulate the effect of a dead zone.

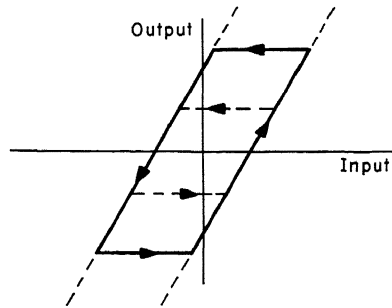


Fig. 14-9-4.

The third common physical nonlinearity is backlash. The representation for this phenomenon is shown in Fig. 14-9-4. Here the similarity to a magnetic hysteresis loop is apparent. A common device with this characteristic is the gear train. Various analog computer circuits have also been devised for simulating backlash.

The final nonlinear characteristic is coulomb friction with the characteristic shown in Fig. 14-9-5. This characteristic illustrates the pres-

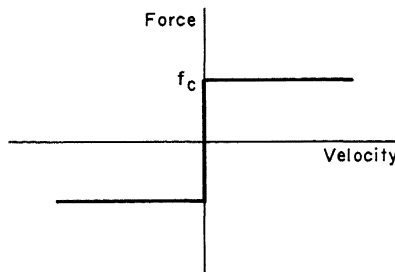


Fig. 14-9-5.

ence of a frictional drag force of constant value that always opposes motion. The analog computer representation of this nonlinear term plus static friction and the other phenomena mentioned above can be found in the literature of the subject.

Inclusion of any of the nonlinearities mentioned above in the analysis or design of a system greatly complicates study by ordinary methods. However an analog computer can handle such problems with only a small amount of additional effort.

PROBLEMS

14-1. Using the force-voltage analogy, construct an electrical circuit model for the mechanical system of Fig. 14-P-1. The weight is initially displaced $x(0)$ inches. Include a suitable analog in your circuit for this initial condition.

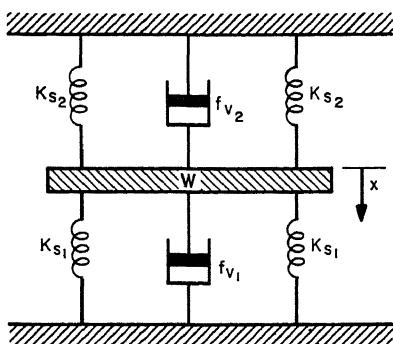


Fig. 14-P-1.

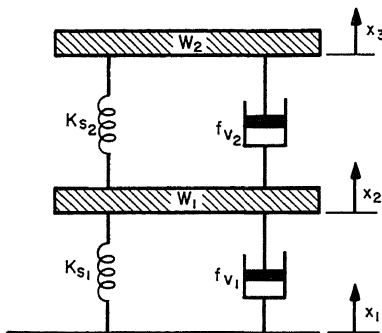


Fig. 14-P-2.

14-2. Repeat problem 14-1 except use the mechanical system of Fig. 14-P-2 and ignore initial conditions. The velocity at x_1 is the input.

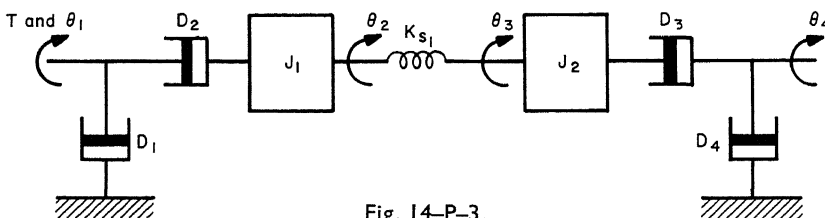


Fig. 14-P-3.

14-3. The rotational system of Fig. 14-P-3 is to be represented by an analogous electrical circuit based on the torque-voltage analogy.

14-4. Check Eqs. 14-2-11 and 14-2-12 dimensionally.

14-5. The mechanical system of Ex. 14-2-1 has the following parameters:

$$K_s = 1000 \text{ lb/in.}$$

$$f_v = 175 \text{ lb/(in./sec)}$$

$$W = 2000 \text{ lbs}$$

$$F = [500 \sin 0.6t]u_{-1}(t)$$

Choose amplitude and time scales for the study of this system using a force-voltage analog circuit. Electrical frequencies should be of the order of 1000 cps, voltages should be of the order of 10 volts and 1/2 watt resistors must be suitable.

14-6. The mechanical system of Ex. 14-2-2 has the following parameters:

$$J_1 = 0.2 \text{ in.-oz-sec}^2$$

$$J_2 = 1.25 \text{ in.-oz-sec}^2$$

$$D_1 = 0.5 \text{ in.-oz-sec}$$

$$D_2 = 2.5 \text{ in.-oz-sec}$$

$$K_s = 5 \text{ in.-oz}$$

$$T = 30 \text{ in.-oz}$$

Scale the circuit of Fig. 14-2-3b for study of repetitive transients by using an oscilloscope and a square-wave generator. The transients should be completed in approximately 0.01 sec or less. Square-wave voltages of up to 50 volts peak-to-peak amplitude are available. R_1 should be at least 1000 ohms. Make estimates based upon attainment of steady-state conditions during each half of the applied square wave and upon neglect of coupling in the system equations.

14-7. Check Eqs. 14-3-8 and 14-3-9 dimensionally.

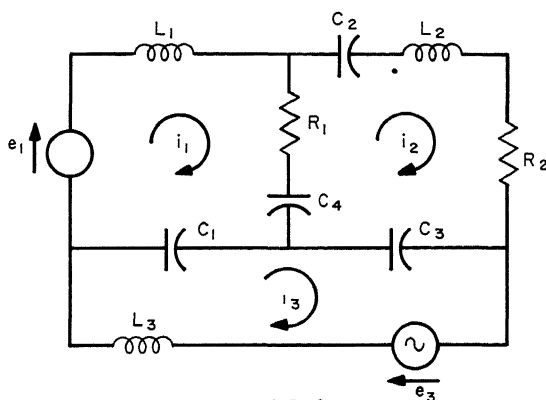


Fig. 14-P-8.

14-8. What translational mechanical system is represented by the electrical circuit of Fig. 14-P-8 using the force-voltage analog.

14-9. The force-current analog is to be used to construct a circuit analogous to the mechanical system of Fig. 14-P-1.

14-10. Repeat Prob. 14-9 for Fig. 14-P-2.

14-11. The mechanical system of Fig. 14-P-2 has

$$W_1 = 5 \text{ lbs}$$

$$W_2 = 100 \text{ lbs}$$

$$K_{s_1} = 210 \text{ lbs/in.}$$

$$K_{s_2} = 200 \text{ lbs/in.}$$

$$f_{v_1} \text{ and } f_{v_2} \text{ variable}$$

The point x_1 is given a velocity of 40 in./sec for 0.1 sec and then stopped. Draw and scale an analogous electrical circuit for the study of damping in this system using

the force-current analogy. Inductances should be 1 henry or less and capacitances 1 microfarad or less. Voltages should be of the order of 50 volts.

14-12. Check Eqs. 14-4-7 and 14-4-8 dimensionally.

14-13. The fluid-handling system of Ex. 14-4-2 has the following parameters:

$$C_{h_1} = 6 \text{ ft}^3 \quad R_{h_1} = 30 \text{ sec/ft}^2$$

$$C_{h_2} = 4 \text{ ft}^3 \quad R_{h_2} = 15 \text{ sec/ft}^2$$

Scale the circuit of Fig. 14-4-2b so that electrical capacitors are 1 microfarad or less and resistors are at least 1 kilohm. Increments in q_1 are of the order of $(1/60)\text{ft}^3/\text{sec}$ and the corresponding current i_1 should not exceed 1 ampere.

14-14. By library reading or otherwise, obtain values for the accuracy to be expected of commercial analog computers. Also, if available, obtain approximate cost of typical computers or components.

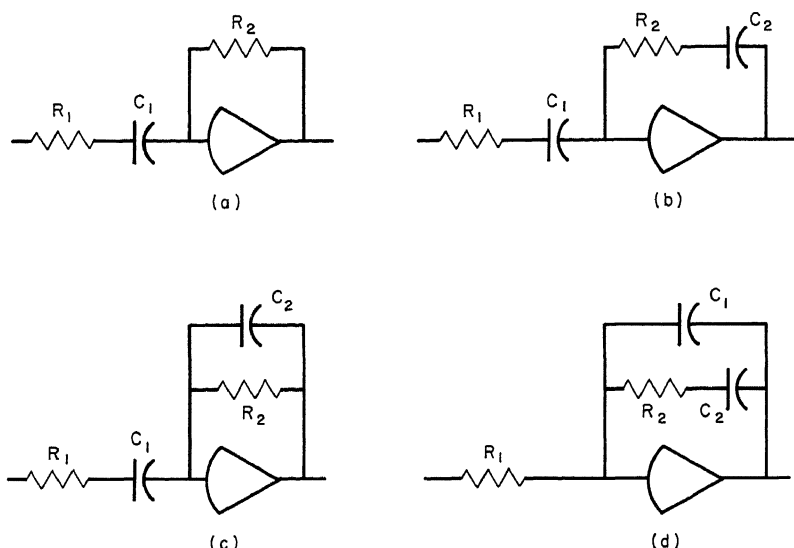


Fig. 14-P-15.

14-15. Derive the operations performed by operational amplifiers using impedances Z_i and Z_f as shown in Fig. 14-P-15.

14-16. Repeat Prob. 14-15 with Z_i and Z_f reversed.

14-17. A slightly modified analysis must be used when paths to ground exist in the input and feedback impedances of operational amplifiers. In Fig. 14-P-17a the voltage e_g is considered zero and $Z_{i_1} = E_i(s)/I_1(s)$ is formed. Similarly with $e_g \approx 0$, $Z_{i_2} = E_o(s)/I_2(s)$ is formed. Finally

$$\frac{E_o(s)}{E_i(s)} = -\frac{Z_{i_2}}{Z_{i_1}}$$

since $i_2 \approx -i_1$. Find the operations performed by the operational amplifiers of Fig. 14-P-17b.

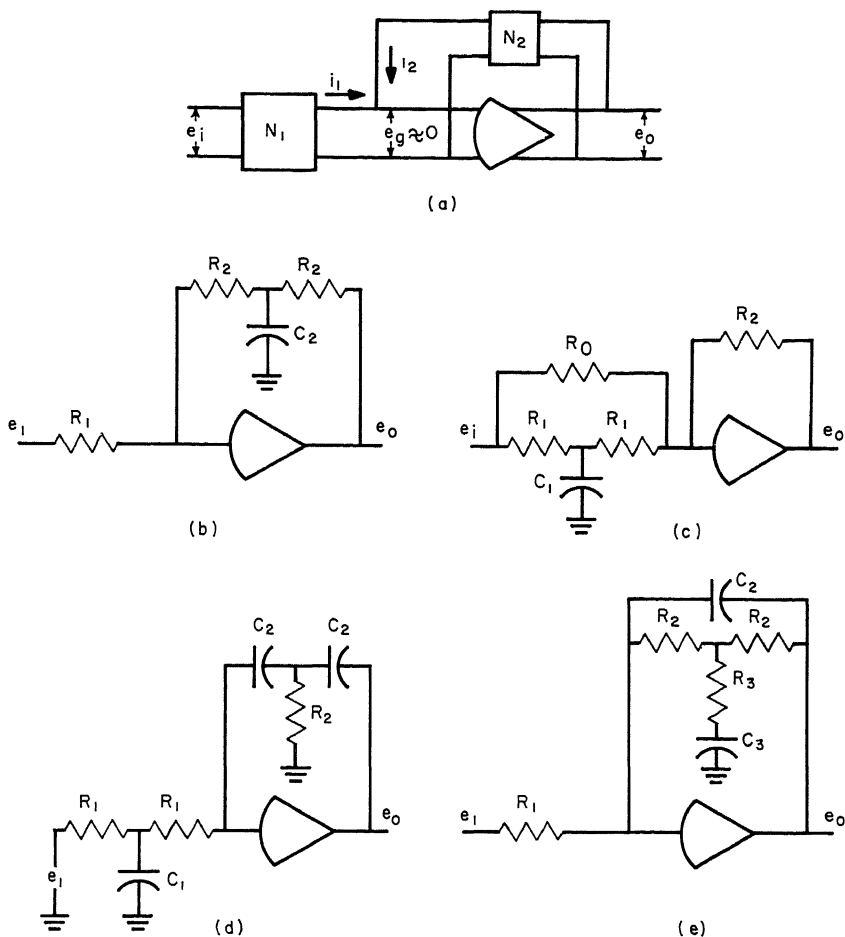


Fig. 14-P-17.

14-18. Repeat Prob. 14-17 with Z_{ti} and Z_{tr} reversed.

14-19. Construct an analog computer block diagram for study of the mechanical system of Fig. 14-P-2. Ignore scaling and use physical variables. Accelerations must be available for recording. Include possible initial positions and velocities.

14-20. Repeat Prob. 4-19 for the mechanical system of Fig. 4-P-3.

14-21. Construct an analog computer diagram for generating $50 \cos(6.28t - 60^\circ)$ volts. Initial conditions are very important.

14-22. Construct an analog computer diagram for study of the fluid-handling system of Ex. 14-4-2. Neglect scaling.

14-23. Construct an analog computer circuit to simulate the closed-loop control system of Fig. 6-7-6b. No differentiation is permitted. Neglect scaling.

14-24. Construct an analog computer circuit to simulate the system shown in the block diagram of Fig. 14-P-24. Neglect scaling.

14-25. Time and amplitude scale an analog computer diagram for study of the system of Prob. 14-5. Maximum voltages must be less than ± 100 volts. The

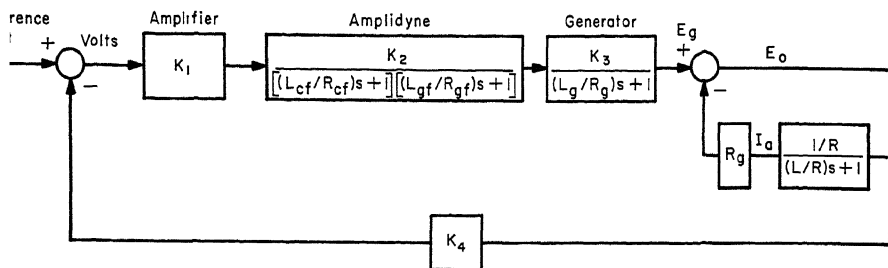


Fig. 14-P-24.

response should reach steady-state in less than 10 seconds. The recorder has a maximum frequency response of 40 cps; however, only 5 cps or less can really be used with good readability of each cycle. (Try a speed-up of 6 times.⁵)

14-26. Use the values of Prob. 14-11 to scale the computer circuit of Problem 14-19. The computer voltages should not exceed ± 100 volts. The optimum natural frequencies are in the neighborhood of 1 cps.⁵

14-27. Program the control system of Problem 6-16 on an analog computer. The computer voltage limits are ± 100 v and natural frequencies should be of the order of 1 cps. Use the simulation method of putting the problem on the computer.⁵

14-28. Repeat Prob. 14-27 for the system of Ex. 6-16-1.⁵

14-29. A galvanometer in a recorder has coupling between its electrical and mechanical circuits as defined by

$$Ri + L \frac{di}{dt} + (BNlb) \frac{d\theta}{dt} = e_s$$

$$J \frac{d^2\theta}{dt^2} + D \frac{d\theta}{dt} + K_s\theta = (BNlb)i$$

where e_s is the voltage applied to the galvanometer, R and L are the galvanometer moving-coil resistance and inductance, B is the magnetic flux density, N is the number of turns on the galvanometer moving coil, and θ is the position of galvanometer needle or moving coil. For one galvanometer the following values are typical:

$$R = 15 \text{ ohms}$$

$$L = 3.5 \text{ mh}$$

$$J = 10 \text{ gm-cm}^2$$

$$K_s = 2.4 \times 10^6 \text{ dyne-cm/rad}$$

The effects on the transient step response of varying $BNlb$ and D of this galvanometer is to be studied on an analog computer. Set this problem up on an analog computer using the computer limitations of Prob. 14-28.⁵

⁵ Either a computer with plug-in impedance elements or one with fixed adder and integrator gains of 1, 4, and 10 may be used. Excessive gains in the plug-in system should not be used (preferably not over 10).

15

SUMMARY

15-1. Practicality. The majority of the theory considered in this text has been developed with highly simplified systems for illustration. Reasonably accurate solutions can be made for many actual systems using second- and third-order equations, but this is not always true. When modern airplane, missile, and processing control systems are considered, the order of the system differential equations may be more nearly fifty or one hundred, and in addition the system will probably be non-linear. Therefore modern computers must be used. Nevertheless the basic theory just studied is needed to provide insight and design information. It would be hopeless to study such large-scale problems without a good understanding of the basic concepts of simple closed-loop systems and the techniques for their analysis and design.

While a certain number of practical limitations of the systems have been pointed out throughout the text, there are many more places in which practical engineering judgment must be used. For example, the synthesis procedure of Chapter 13 and, for that matter, the frequency response methods of Chapters 7, 8, and 9 ignore the physical limitations of the system motor or actuator. Perhaps the compensation scheme would require the actuator to operate linearly with 100 times rated voltage. This solution would be absurd. Poles and zeros simply do not tell the whole story. In other cases a solely theoretical design might call for 1 farad capacitors or 1000 megohm resistors, which are unrealistic because of size and price, not to mention the excessive noise introduced by such a resistor. The practical side of system design, unfortunately, or perhaps fortunately, is often learned by experience. However, additional study can prevent many expensive mistakes.

15-2. General Summary. While studying this text the student has asked himself, and probably his instructor, why so many techniques are needed and where they were leading. By now these questions may have been answered, but in any event a final detached look at both the forest and its trees seems in order. A control system course brings together for the first time the theory given in earlier courses in different areas. Elec-

trical machinery and electronics come face to face. Transient and steady-state analysis methods are applied to many different types of systems. There is a broadening of the concepts of the dynamics of mechanical and hydraulic systems. All of the foregoing is encompassed by one or two courses. This broadness of scope causes many students to grasp for textbooks they thought were no longer needed, and to search out handbooks with various systems of units. Undoubtedly such a unifying course would be worth while even if it added no new approaches; but in reality one of the main objectives is to put some broad new tools for design and analysis into the engineer's bag of tricks. Certainly the terms "steady state" and "transient" are given broader significance. These terms cut across the engineering fields and it is hoped that the reader now sees them in a far more general light. A closed-loop system—a system in which the output is compared with the input and the difference is used to actuate the system—is at least a fascinating phenomenon. The first laboratory contact with such a phenomenon is a memorable experience for many students. From the beginning of closed-loop analysis in Chapter 6, through Chapter 14, an attempt is made to add methods of transient and steady-state analysis that will not leave the student dependent on grinding out transforms to get answers to various linear problems. The more tools an engineer has at his disposal, the better is his ability to analyze and design systems that must meet diverse specifications. Chapters 7, 8, and 9 add the Nyquist criterion and the Bode steady-state frequency-response methods that facilitate designs for specified M_m , BW , phase margin, etc. Contrasted with these are the pole-zero, root-locus, and direct-synthesis methods that strike at the problem of time (and to a certain extent frequency) response when specifications are in terms of overshoot, rise time, BW , etc. The methods presented thus supplement each other. The relations between the system frequency characteristics and the time response are presented in more detail in advanced texts and in the literature. In some cases the previous chapters present only the mechanics of methods with general plausibility arguments. For example, the proofs of the Nyquist criterion and the Bode phase integral relating phase angle to the slope of gain curves are omitted. These items point to the more advanced areas lying beyond the material presented in this text.

15-3. Extension of Control System Analysis. Various additional topics that need to be studied by the control system student are pointed out in this article in order to provide a broader picture of the students situation at this point. Probably of most immediate importance is work in the area of minor loops and non-unity feedback systems. The benefits of such schemes over series compensation must be considered. The effects of load disturbances, such as added torques, directly enter this picture. Multiple-input systems and systems with noise or stray signals also need study. This text considers only superficially systems that oper-

ate on carrier frequencies. These so-called a-c systems require special techniques for compensation and detailed study. Beyond these areas is the comparatively new field of statistical design. This area deals with design, using statistical properties of signals, and investigates such things as minimizing the rms error and various other error criteria. Obviously there is a wealth of linear material still to be studied, to say nothing of nonlinear items.

Much research is being concentrated on the area of nonlinear controls. Studies include optimizing criteria, methods of design to utilize nonlinearities, and systems that modify their own parameters to adapt to environmental changes in order to give optimum performance. In short, the goal is systems that possess certain capabilities far greater than those of their human operators.

A slightly different approach is an area termed sampled-data control systems. Such systems work with signals that are discrete samples rather than continuous variables. The use of digital computers in closed-loop sampled-data systems poses problems in this field. Certainly there is no dearth of material to study and no limit to what the reader may someday encounter, or perhaps design, in the area of control systems.

APPENDIX *

* Reprinted from G. V. Lago and D. L. Waideich, *Transients in Electrical Circuits*, The Ronald Press Co., New York, 1958.

LAPLACE TRANSFORM FORMULAS

	$F(s)$	$f(t)$
1	$\frac{1}{s}$	$u_{-1}(t)$
2	$\frac{1}{s^n}$, n is a positive integer	$\frac{t^{n-1}}{(n-1)!}$
3	$\frac{1}{s+a}$	e^{-at}
4	$\frac{1}{(s+a)^n}$, n is a positive integer	$\frac{t^{n-1} e^{-at}}{(n-1)!}$
5	$\frac{\omega}{s^2 + \omega^2}$	$\sin \omega t$
6	$\frac{s}{s^2 + \omega^2}$	$\cos \omega t$
7	$\frac{b}{s^2 - b^2}$	$\sinh bt$
8	$\frac{s}{s^2 - b^2}$	$\cosh bt$
9	$\frac{\omega}{(s+a)^2 + \omega^2}$	$e^{-at} \sin \omega t$
10	$\frac{s+a}{(s+a)^2 + \omega^2}$	$e^{-at} \cos \omega t$
11	$\frac{b}{(s+a)^2 - b^2}$	$e^{-at} \sinh bt$

12	$\frac{s+a}{(s+a)^2-b^2}$	$\epsilon^{-at} \cosh bt$
13	$\frac{a_1 s + a_0}{(s+a)^2 + \omega^2}$	$A \epsilon^{-at} \sin (\omega t + \alpha)$ where $A/\alpha = \left[\frac{a_0 - a_1(a-j\omega)}{\omega} \right]$
14	$\frac{a_1 s + a_0}{(s+a)(s+b)}$	$\left(\frac{a_0 - a_1 a}{b-a} \right) \epsilon^{-at} + \left(\frac{a_0 - b a_1}{a-b} \right) \epsilon^{-bt}$
15	$\frac{a_1 s + a_0}{(s+a)^2}$	$[a_1 + (a_0 - a_1 a)t] \epsilon^{-at}$
16	$\frac{a_2 s^2 + a_1 s + a_0}{(s+a)(s+b)(s+c)}$	$\frac{a_2 a^2 - a_1 a + a_0}{(b-a)(c-a)} \epsilon^{-at} + \frac{a_2 b^2 - a_1 b + a_0}{(a-b)(c-b)} \epsilon^{-bt} + \frac{a_2 c^2 - a_1 c + a_0}{(a-c)(b-c)} \epsilon^{-ct}$
17	$\frac{a_2 s^2 + a_1 s + a_0}{(s+a)^2(s+b)}$	$\left[\frac{b(a_1 - 2a_2 a) + a_2 a^2 - a_0}{(b-a)^2} \right] \epsilon^{-at} + \left(\frac{a_2 a^2 - a_1 a + a_0}{b-a} \right) t \epsilon^{-at} + \left[\frac{a_2 b^2 - a_1 b + a_0}{(b-a)^2} \right] \epsilon^{-bt}$
18	$\frac{a_2 s^2 + a_1 s + a_0}{(s+a)^3}$	$a_2 \epsilon^{-at} + (a_1 - 2a_2 a)t \epsilon^{-at} + \frac{1}{2}(a_2 a^2 - a_1 a + a_0)t^2 \epsilon^{-at}$
19	$\frac{a_2 s^2 + a_1 s + a_0}{[(s+a)^2 + \omega^2](s+b)}$	$A \epsilon^{-at} \sin (\omega t + \alpha) + \frac{a_0 - a_1 b + a_2 b^2}{(a-b)^2 + \omega^2} \epsilon^{-bt}$ where $A/\alpha = \frac{a_0 - a_1(a-j\omega) + a_2(a-j\omega)^2}{\omega(b-a+j\omega)}$
20	$\frac{a_3 s^3 + a_2 s^2 + a_1 s + a_0}{(s+a)(s+b)(s+c)(s+d)}$	$\frac{a_0 - a_1 a + a_2 a^2 - a_3 a^3}{(b-a)(c-a)(d-a)} \epsilon^{-at} + \frac{a_0 - a_1 b + a_2 b^2 - a_3 b^3}{(a-b)(c-b)(d-b)} \epsilon^{-bt}$ + $\frac{a_0 - a_1 c + a_2 c^2 - a_3 c^3}{(a-c)(b-c)(d-c)} \epsilon^{-ct} + \frac{a_0 - a_1 d + a_2 d^2 - a_3 d^3}{(a-d)(b-d)(c-d)} \epsilon^{-dt}$

LAPLACE TRANSFORM FORMULAS (Continued)

	$F(s)$	$f(t)$
21	$\frac{a_3 s^3 + a_2 s^2 + a_1 s + a_0}{(s+a)^2(s+b)(s+c)}$	$\frac{a_0 - a_1 a + a_2 a^2 - a_3 a^3}{(b-a)(c-a)} t e^{-at} + \left[\frac{a_1 2 a_2 a + 3 a_3 a^2}{(b-a)(c-a)} - \frac{(a_0 - a_1 a + a_2 a^2 - a_3 a^3)(b+c-2a)}{(b-a)^2(c-a)^2} \right] e^{-at}$ $+ \frac{a_0 - a_1 b + a_2 b^2 - a_3 b^3}{(a-b)^2(c-b)} e^{-bt} + \frac{a_0 - a_1 c + a_2 c^2 - a_3 c^3}{(a-c)^2(b-c)} e^{-ct}$
22	$\frac{a_3 s^3 + a_2 s^2 + a_1 s + a_0}{(s+a)^2(s+b)^2}$	$\frac{(a_0 - a_1 a + a_2 a^2 - a_3 a^3)}{(b-a)^2} t e^{-at} + \frac{(a_0 - a_1 b + a_2 b^2 - a_3 b^3)}{(a-b)^2} t e^{-bt}$ $+ \left\{ \frac{a_1 - 2 a_2 a + 3 a_3 a^2}{(b-a)^2} - \frac{2(a_0 - a_1 a + a_2 a^2 - a_3 a^3)}{2(a-b)^3} \right\} e^{-at}$ $+ \left\{ \frac{a_1 - 2 a_2 b + 3 a_3 b^2}{(a-b)^2} - \frac{2(a_0 - a_1 b + a_2 b^2 - a_3 b^3)}{(a-b)^3} \right\} e^{-bt}$
23	$\frac{a_3 s^3 + a_2 s^2 + a_1 s + a_0}{(s+a)^3(s+b)}$	$\frac{a_0 - a_1 a + a_2 a^2 - a_3 a^3}{2(b-a)} t^2 e^{-at} + \frac{a_0 - a_1 b + a_2 b^2 - a_3 b^3}{(a-b)^3} e^{-bt}$ $+ \left[\frac{a_1 - 2 a_2 a + 3 a_3 a^2}{b-a} - \frac{a_0 - a_1 a + a_2 a^2 - a_3 a^3}{(b-a)^2} \right] t e^{-at}$ $+ \left[\frac{a_2 - 3 a_3 a}{b-a} - \frac{a_1 - 2 a_2 a + 3 a_3 a^2}{(b-a)^2} + \frac{a_0 - a_1 a + a_2 a^2 - a_3 a^3}{2(b-a)^3} \right] e^{-at}$
24	$\frac{a_3 s^3 + a_2 s^2 + a_1 s + a_0}{(s+a)^4}$	$\frac{1}{6}(a_0 - a_1 a + a_2 a^2 - a_3 a^3) t^3 e^{-at} + \frac{1}{2}(a_1 - 2 a_2 a + 3 a_3 a^2) t^2 e^{-at} + (a_2 - 3 a_3 a) t e^{-at} + a_3 e^{-at}$
25	$\frac{a_3 s^3 + a_2 s^2 + a_1 s + a_0}{[(s+a)^2 + \omega^2](s+b)(s+c)}$	$A e^{-at} \sin(\omega t + \alpha) + \frac{a_0 - a_1 b + a_2 b^2 - a_3 b^3}{[(a-b)^2 + \omega^2](c-b)} e^{-bt} + \frac{a_0 - a_1 c + a_2 c^2 - a_3 c^3}{[(a-c)^2 + \omega^2](b-c)} e^{-ct}$ where $A/\alpha = \frac{a_0 - a_1(a-j\omega) + a_2(a-j\omega)^2 - a_3(a-j\omega)^3}{\omega(b-a+j\omega)(c-a+j\omega)}$

26	$\frac{a_3 s^3 + a_2 s^2 + a_1 s + a_0}{[(s+a)^2 + \omega^2][(s+b)^2 + \sigma^2]}$	$A \epsilon^{-at} \sin(\omega t + \alpha) + \frac{a_0 - a_1 b + a_2 b^2 - a_3 b^3}{[(a-b)^2 + \omega^2]} t \epsilon^{-bt} \\ + \left\{ \frac{a_1 - 2 a_2 b + 3 a_3 b^2}{[(a-b)^2 + \omega^2]} - 2 \frac{(a_0 - a_1 b + a_2 b^2 - a_3 b^3)(a-b)}{[(a-b)^2 + \omega^2]^2} \right\} \\ \text{where } A/\alpha = \frac{a_0 - a_1(a-j\omega) + a_2(a-j\omega)^2 - a_3(a-j\omega)^3}{\omega(b-a+j\omega)^2}$
27	$\frac{a_3 s^3 + a_2 s^2 + a_1 s + a_0}{[(s+a)^2 + \omega^2][(s+b)^2 + \sigma^2]}$	$A \epsilon^{-at} \sin(\omega t + \alpha) + B \epsilon^{-bt} \sin(\sigma t + \beta) \\ \text{where } A/\alpha = \frac{a_0 - a_1(a-j\omega) + a_2(a-j\omega)^2 - a_3(a-j\omega)^3}{\omega[(b-a+j\omega)^2 + \sigma^2]} \\ B/\beta = \frac{a_0 - a_1(b-j\sigma) + a_2(b-j\sigma)^2 - a_3(b-j\sigma)^3}{\sigma[(a-b+j\sigma)^2 + \omega^2]}$
28	$\frac{a_3 s^3 + a_2 s^2 + a_1 s + a_0}{[(s+a)^2 + \omega^2]^2}$	$- \frac{A t \epsilon^{-at}}{2 \omega^2} \cos(\omega t + \alpha) + 2 B \epsilon^{-at} \cos(\omega t + \beta) \\ \text{where } A/\alpha = a_0 - a_1(a-j\omega) + a_2(a-j\omega)^2 + a_3(a-j\omega)^3 \\ B/\beta = \frac{a_1 - 2 a_2(a-j\omega) + 3 a_3(a-j\omega)^2 - 2[a_0 - a_1(a-j\omega) + a_2(a-j\omega)^2 - a_3(a-j\omega)^3]}{(2 j \omega)^2}$
29	$\frac{a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_0}{(s^2 + \omega^2)(s+a)(s+b)(s+c)}$	$\frac{a_0 - a_1 a + a_2 a^2 - a_3 a^3 + a_4 a^4}{(a^2 + \omega^2)(b-a)(c-a)} \epsilon^{-at} + \frac{a_0 - a_1 b + a_2 b^2 - a_3 b^3 + a_4 b^4}{(b^2 + \omega^2)(a-b)(c-b)} \epsilon^{-bt} \\ + \frac{a_0 - a_1 c + a_2 c^2 - a_3 c^3 + a_4 c^4}{(c^2 + \omega^2)(a-c)(b-c)} \epsilon^{-ct} + A \sin(\omega t + \alpha) \\ \text{where } A/\alpha = \frac{a_0 + a_1(j\omega) + a_2(j\omega)^2 + a_3(j\omega)^3 + a_4(j\omega)^4}{\omega(j\omega + a)(j\omega + b)(j\omega + c)}$
30	$\frac{a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_0}{(s^2 + \omega^2)(s+a)^2(s+b)}$	$\frac{a_0 - a_1 a + a_2 a^2 - a_3 a^3 + a_4 a^4}{(a^2 + \omega^2)(b-a)} t \epsilon^{-at} + \frac{a_0 - a_1 b + a_2 b^2 - a_3 b^3 + a_4 b^4}{(b^2 + \omega^2)(a-b)^2} \epsilon^{-bt} \\ + \left[\frac{a_1 - 2 a_2 a + 3 a_3 a^2 - 4 a_4 a^3}{(a^2 + \omega^2)(b-a)} - \frac{(a_0 - a_1 a + a_2 a^2 - a_3 a^3 + a_4 a^4)(\omega^2 - 2 ab + 3 a^2)}{(a^2 + \omega^2)^2(b-a)^2} \right] \epsilon^{-at} \\ + A \sin(\omega t + \alpha) \text{ where } A/\alpha = \frac{a_0 + a_1(j\omega) + a_2(j\omega)^2 + a_3(j\omega)^3 + a_4(j\omega)^4}{\omega(j\omega + a)^2(j\omega + b)}$

LAPLACE TRANSFORM FORMULAS (Continued)

	$F(s)$	$f(t)$
31	$\frac{a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_0}{(s^2 + \omega^2)(s + a)^3}$	$A \sin(\omega t + \alpha) + \frac{a_0 - a_1 a + a_2 a^2 - a_3 a^3 + a_4 a^4}{2(a^2 + \omega^2)} t^2 e^{-at}$ $+ \left[\frac{a_1 - 2a_2 a + 3a_3 a^2 - 4a_4 a^3}{a^2 + \omega^2} + \frac{2a(a_0 - a_1 a + a_2 a^2 - a_3 a^3 + a_4 a^4)}{(a^2 + \omega^2)^2} \right] t e^{-at}$ $+ \left[\frac{2a_2 - 6a_3 a + 12a_4 a^2}{(a^2 + \omega^2)} + \frac{4a(a_1 - 2a_2 a + 3a_3 a^2 - 4a_4 a^3)}{(a^2 + \omega^2)^2} \right. \\ \left. + \frac{(6a^2 - 2\omega^2)(a_0 - a_1 a + a_2 a^2 - a_3 a^3 + a_4 a^4)}{(a^2 + \omega^2)^3} \right] e^{-at}$ where $A/\alpha = \frac{a_0 + a_1(j\omega) + a_2(j\omega)^2 + a_3(j\omega)^3 + a_4(j\omega)^4}{\omega(j\omega + a)^3}$
32	$\frac{a_4 s^4 + a_3 s^3 + a_2 s^2 + a_1 s + a_0}{(s^2 + \omega^2)[(s + a)^2 + \sigma^2](s + b)}$	$A \sin(\omega t + \alpha) + B e^{-at} \sin(\sigma t + \beta) + \frac{a_0 - a_1 b + a_2 b^2 - a_3 b^3 + a_4 b^4}{(b^2 + \omega^2)[(a - b)^2 + \sigma^2]} e^{-bt}$ where $A/\alpha = \frac{a_0 + a_1(j\omega) + a_2(j\omega)^2 + a_3(j\omega)^3 + a_4(j\omega)^4}{\omega[(j\omega + a)^2 + \sigma^2](j\omega + b)}$ $B/\beta = \frac{a_0 - a_1(a - j\sigma) + a_2(a - j\sigma)^2 - a_3(a - j\sigma)^3 + a_4(a - j\sigma)^4}{\sigma[(a - j\sigma)^2 + \omega^2](b - a + j\sigma)}$

INDEX

- α , loci of constant, 242-247
- Acceleration constants
 - dynamic, 346
 - static, 186
- Actuating signal, defined, 5
- Adaptive systems, 493
- Adder, operational amplifier, 469
- Admittance, 57, 65
- Amplifiers
 - amplidyne, 107
 - operational, 468
 - vacuum tube, 78
- Amplitude scaling, 459, 463
 - in analog computers, 477
- Analog computer, 468-485
 - amplitude scaling, 477
 - nonlinear operation, 486
 - operational amplifier, 468
 - problem set-up, 470
 - time scaling, 475
- Analogs
 - hydraulic-electric, 466
 - mechanical-electrical, 457
 - force-current, 462
 - force-voltage, 457
 - torque-current, 464
 - torque-voltage, 461
- Analytic continuation, 39
- Analytic functions, 36
 - Cauchy-Riemann conditions, 37
- Angelo, E. J., 79, 81
- Aseltine, John A., 371, 377
- Asymptotes
 - for Bode diagram, 288-310
 - for root-locus, 357
- Attenuation diagram; *see* Bode diagram
- Band elimination filter, 267, 343
- Band pass filter, 343
- Bandwidth, 267, 387
 - defined, 154
 - used in direct synthesis, 437-454
- Block diagram
 - defined, 69
 - for electric circuit, 69-81
 - for electromechanical system, 104-123
 - for mechanical system, 94-104
 - reduction of system to, 161-166
 - transformations, 154-161
- Bode, H. W., 380
- Bode diagram, 282-310, 315
 - attenuation plots, 283-293
 - break frequencies, 289-293
 - gain adjustment, 289, 298
 - lag compensation, 301, 305
 - lag-lead compensation, 308
 - lead compensation, 302, 305
 - phase curves, 293-295
 - quadratic factor, 292, 295
 - stability in terms of, 297
- Bower and Schultheiss, 301
- Break frequencies, 289-293
- Brune, Otto, 415
- Bruns and Saunders, 91
- Bulk modulus, 129
- Butterworth functions, 339
- Capacitor, 52, 57
- Capillary-bellows system, 146
- Carrier frequencies, 493
 - envelope transfer function, 121
- Cauchy-Riemann conditions, 37
- Cauer, Wilhelm, 415
- Cauer's forms
 - first form, 402, 410
 - second form, 404, 411
- Characteristic polynomial, 183, 205
 - determining zeros of, 184, 205, 217
- Chu, Yaohan, 392
- Closed-loop control system introduced, 3, 7
- Coefficient potentiometer, 469
- Compensation, 254-272, 436-454
 - discussion of, 271
 - for system type number, 267
 - parallel, 255, 269-271
 - series, 256-276
 - lag-lead network, 263-276
 - lag network, 260-263
 - lead network, 256-260
 - more general networks, 436-454
- Complementary solution, 13
- Complete solution; *see* General solution
- Computer; *see also* Analog computer
 - units, 478
- Conformal mapping, 218-221
- Constant multipliers, 469
- Continued-fraction expansion, 403-411

- Convergence, abscissa of, 24-26
- Convolution, 22, 318-341
 - graphical interpretation, 321-324
 - integral, 22, 318-321
 - uses, 324-341
- Coulomb friction, 485
- "Cover-up" technique, 41-43
- Critical point, 220
- Curve follower, 9
- Cybernetics, 6
- D-c generator, 8, 10, 104
- D-c motor, 10
 - armature controlled, 112
 - field controlled, 117
- Damping, 90
- Damping ratio, 168
- Dead-time delay operator, 26
- Dead zone, 484
- Decibels, 282, 283-310
 - per decade, 284-293
 - per octave, 284-293
- Delay time, 153
- Descartes' rule of signs, 206
- Differential equations, 11
- Differentiating circuit, 273
- Differentiation in generalized sense, 30-33, 63, 320-327
- Direct root-locus; *see* Root-locus
- Direct synthesis methods, 436-454
- Direct transform; *see* Laplace transform; Fourier transform
- Discontinuity, 30-33, 63, 320-327
- Disturbance, 4, 7
- Dominant poles, 186, 194
- Driving point impedance
 - LC, 397
 - RC, 398, 412
 - RL, 399
- Eckman, Donald P., 130
- Electric machines, 8, 10, 112-117
- Electromechanical systems, 104-123
- Elgerd and Stephens, 387
- Energy relations, 90, 92, 130
- Envelope transfer function, 121
- Equality of form, 458, 462
- Error coefficients
 - dynamic, K'_p , K'_v , K'_a , 346, 348
 - static, K_p , K_v , K_a , 186, 345
- Error rate damping, 178
- Error series, 344
- Error signal, 5
- Evans, W. R., 352
- Even functions, 206, 399
- Expansion theorem, 43
- Exponential order, 16
- $F(s)$, 16
 - discussion of, 36
 - partial fraction expansion, 40
 - product of two, 22
- Factoring
 - extended Horner's method, 215
 - Lin's method, 184, 206
- Faraday's law, 462
- Feedback
 - from input shaft, 180
 - from output shaft, 177
- Feedback compensation, 255, 269-271
- Feedback path, 5
- Final value theorem, 45
- Fluid flow laws
 - laminar flow, 130
 - nonlaminar flow, 129
- Force-current analog, 462
- Forced response, 13
- Force-voltage analog, 457
- Foster's forms
 - first, 399, 406, 416
 - second, 401, 409
- Fourier series, 45
- Fourier transform
 - direct, 315
 - inverse, 318
- Free body diagram, 93
- Frequency response, 13, 63, 96, 154, 190, 315
 - from poles and zeros, 195
- Frequency scaling, 175
- Friction
 - coulomb, 485
 - viscous, 88, 458
- $G(s)$, 68, 156
- Gain, 233, 236, 237, 239, 241, 289
- Gain adjustment, 379
 - in terms of Bode diagram, 298
 - in terms of Nyquist diagram, 252
 - in terms of root-locus, 378
 - setting for M_m , 252, 254, 258, 262
- Gain constants
 - dynamic, K'_p , K'_v , K'_a , 346, 348
 - static, K_p , K_v , K_a , 186, 345
- Gain margin, 297, 300, 302, 305
- Gain of networks, 424, 426, 429
 - adjusting, 430
- Gardner and Barnes, 462
- Gas constant, universal, 144
- Gear trains, 101-104
- Generalized curves, 169, 172, 173
- Generalized equation, 168
- General solution, 13
- Gibson and Tuteur, 111, 120, 138

- $H(s)$, 156
- Hausenbauer and Lago, 387, 437, 443-454
- Heaviside expansion theorem, 43
- High-pass filter, 267, 343
- Hurwitz polynomials, 207-218, 399
- Hydraulic capacitance, 130, 132
- Hydraulic-electric analogs, 466
- Hydraulic resistance, 130, 132
- Hydraulic systems, 11, 128-140, 164
 - pump controlled motor, 138
 - valves and actuator, 134
- Impedance, 57, 65
- Impulse family of functions, 28-33, 62-63, 76-78, 397-398
- Impulse functions used for initial conditions, 76-78
- Incremental circuit, 80, 81
- Incremental plate resistance, 79
- Incremental technique for linearization, 78, 131, 142, 145-147
- Incremental transconductance, 79
- Inductance, 13
- Inductor, 57, 65
- Initial conditions, 76, 95, 197
 - added to operational approach, 76
 - in examples, 53, 55
 - in transform of derivatives, 18, 19
- Initial-value theorem, 45
- Input signals, 152
 - sinusoidal, 13, 63, 96, 154, 190, 195, 315
 - unit-impulse, 166
 - unit-ramp, 177
 - unit-step, 170
- Integral compensation, 181
- Integrating circuit, 276
- Integrator, operational amplifier, 469
- Inverse complex-plane diagram, 246
- Inverse Fourier transform, 318
- Inverse Laplace transform, 17, 39, 317
- Inverse root-locus; *see* Root-locus
- Iron-core inductor, 14
- Johnson, Clarence, 483
- K_a , 186, 345
- K_p , 186, 345
- K_v , 186, 252-262, 345, 383-388, 438-454
- Kirchhoff's current law, 55, 462
- Kirchhoff's voltage law, 53, 458
- Korn and Korn, 483
- Lag compensation
 - Bode diagram, 301
 - Nichol's diagram, 301
- Nyquist diagram, 262
- root-locus, 384
- Lag-lead compensation
 - Bode diagram, 305-310
 - Nichol's diagram, 305-310
 - Nyquist diagram, 266
- Lag-lead network, 263-276
 - poles and zeros of, 263
- Lag network, 260-263
 - comparison of transient and frequency terminology, 275
 - poles and zeros of, 261
- Lag unit, operational amplifier, 469
- Lane, L. Jubin, 478
- Laplace transform, 16, 316, 317
 - derivative of, 23
 - direct, 16
 - inverse, 17, 39
 - table, 496-500
 - theorems
 - dead-time delay, 21
 - derivative, 17-19
 - final-value, 45
 - initial-value, 45-47
 - integral, 19
- Lead compensation
 - Bode diagram, 302
 - Nichol's diagram, 302
 - Nyquist diagram, 257
 - root-locus, 379
- Lead network, 256-260
 - comparison of transient and frequency terminology, 272
 - poles and zeros of, 257
- Linear differential equations, 11
- Lin's method of factoring, 184, 206
- Liquid level control, 11
- Loading effect, 74
- Location of zeros, 206
- Low-pass filters, 267, 343
- μ , 79, 80
- M -circles, 242-247
- M_m , 154, 190, 301, 315
- Magnetic saturation, 484
- Mapping
 - conformal, 218-221
 - s -plane to $[1 + H(s)G(s)]$ -plane, 223-225
 - s -plane to $[H(s)G(s)]$ -plane, 225-234
- Mass, 88, 91-93
- Maximally flat functions, 338
- Mechanical differential, 10, 104
- Mechanical systems, 88-105
- Mixing point, 5-8
- Moment of inertia, 88-91
- Motor, d-c, 10, 112, 117

- Motor, hydraulic, 138
- Multiple-loop systems, 159
- Multiplication of two transforms, 22, 318-341
- N*-circles, 244-247
- n*th order differential equation, 12
- Natural frequency, 154, 167, 190
- Natural response, 13
- Networks
 - lag, 260-263
 - lag-lead, 263-276
 - lead, 256-260
- Network synthesis, 397-432, 448, 452
- Newton's second law, 93
- Nichol's chart, 298-310
- Nonlinear operation, 483
- Nonlinear systems, 13
- Nonminimum phase, 341
- Normalized velocity constant, 439
- Normalizing, 257, 261, 265, 438
- Nyquist diagram, 221-278, 282, 315, 380
 - from experimental data, 272
 - relative stability, 230-232, 240-245
 - type 0 system, 235
 - type 1 system, 227-234, 236, 240
 - type 2 system, 237-239
 - type 5 system, 240
- ω_m , 154, 190
- ω_n , 167
- Odd functions, 206, 399
- Open-loop control systems, 3, 7
- Operational amplifier, 468
- Operational impedance, 57, 65
- Overshoot, 153
- Parallel compensation, 255, 269-271
- Partial-fraction expansion, 40
- Particular solution, 12, 63
- Phase diagrams; *see* Bode diagram
- Phase lag network; *see* Lag network
- Phase lead network; *see* Lead network
- Phase margin, 241, 297, 300, 302, 305, 309, 380
- Phasor impedance from operational impedance, 65
- Phasor transforms from operational transforms, 65
- Pick-off point, 156
- Pneumatic capacitance, 142, 144
- Pneumatic flow equations
 - smooth flow through tubes and capillaries, 140
 - subsonic, 140
 - supersonic, 141
- Pneumatic resistance, 142
- Pneumatic systems, 140-148
- Pole
 - definition, 37
 - geometric interpretation, 38
- Pole-zero
 - determination of frequency response, 195
 - determination of transient response, 192
 - diagram, 38, 105, 106, 108, 114, 118, 122, 170, 192, 195, 342, 397, 398, 399, 415
- Position constant
 - dynamic, 346
 - in terms of poles and zeros, 347
 - static, 186
- Potts and Thaler, 306
- Power relations, 90, 93, 130, 139
- Propeller milling machine, 10
- Quadratic form, 417
- RC* networks, 396, 406-432; *see also* Lag network; Lag-lead network; Lead network
- Re*, 64
- Reactance functions, 207
- Real time; *see* Time scaling
- Reference input; *see* Input signals
- Relative stability, 215, 240-247
- Residue, 40-42
 - graphical interpretation of, 195
- Residue condition, 417
- Resistor, 14, 52, 57
- Reynolds number, 129
- Rise time, 153
- RL* networks, 399
- Root-locus, 352-392
 - basis for direct, 352
 - basis for inverse, 353
 - comments on direct, 369
 - comments on inverse, 357
 - examples for direct, 375-371
 - examples for inverse, 371-378
 - generalized compensation, 385
 - lag compensation, 384
 - lead compensation, 379
 - non-unity feedback, 391
 - rules for direct, 354-369
 - rules for inverse, 371-377
- Roots; *see* Factoring
- Rotational mechanical systems, 99
- Routh criterion, 206-218
- Routh table, 206-218, 404
- s*, complex variable, 16, 36
- Sampled-data systems, 493

- Saturation or limiting, 484
- Scaling, 459, 463, 467
- Sectionally continuous, 16
- Series
 - error, 344
 - Fourier, 45
- Series compensation, *see* Compensation
- Servomotor, 112
- Settling time, 153
- Singular point, 37
- Sinusoidal steady-state response, 63-66,
 - 96, 154, 190, 195, 315-317
 - from poles and zeros, 195
- Speed-torque curves, 115, 116
- Spirule, 369
- Spring constant, 88, 91
- Spring-mass-friction system, 88
- Stability
 - Bode diagram, 283-310
 - factoring, 205
 - Nyquist diagram, 221-247
 - root-locus diagram, 352-392
 - Routh criterion, 221-247
- Static error coefficients, 186
- Statistical design, 493
- Steady-state solution, 13
- Summing integrator, 469
- Summing point, 155
- Synchro units
 - generator, 109
 - motor, 109
 - transformer, 110
- Synthesis; *see also* Network synthesis
 - direct, 436-454
 - 2 poles, no zeros, 437
 - 2 poles, one zero, 439
 - 3 poles, one zero, 439
- Tachometer, 177, 180
- Temperature control, 7
- Thaler and Stein, 112
- Thermostat, 7
- Time scaling, 459, 463, 475
- Transfer admittance, 70, 414
- Transfer functions, 67-76
 - closed-loop systems, 157
 - network in tandem, 72-74
 - RC* networks, 421-431
 - vacuum tube circuits, 78
- Transfer impedance, 68-69, 413, 415
- Transform pairs, table of, 496-500
- Transient response, 13, 166, 183-189,
 - 328-337
 - discussion of, 189
 - from poles and zeros, 337
 - higher-order systems, 183-186
 - simple systems, 328
- Truxal, John G., 110, 436
- Two-phase a-c servomotor, 120-123
- Two-terminal-pair network, 412
- Type number, 187-189, 234, 267, 344
 - and Nyquist diagram, 234
- Unit-impulse function
 - driving function, 62
 - in convolution, 323
 - transform of, 28
- Unit-step function
 - driving function, 61
 - in convolution, 318
 - transform of, 24
- Universal curves, 169, 172, 173, 191
- Unstable systems; *see* Stability
- Vacuum tube circuits, 78-81
- Velocity constant
 - dynamic, 346
 - in terms of poles and zeros, 347
 - static, 186
- Viscosity, absolute, 129
- Viscous friction, 88, 458
 - effect due to back emf of motor, 162-163
- Voltage-force analog, 457
- Voltage regulation, 8
- Wiener, N. W., 6
- Zaborsky, John, 371, 387
- Zero, 37
- Zeros of polynomials; *see* Factoring

W
2992